Selection of Binders and Inhibitors from Designed Ankyrin Repeat Protein Libraries

DISSERTATION

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

UNIVERSITÄT ZÜRICH

von

Patrick Amstutz

von Thalwil / ZH

Promotionskomitee

Prof. Dr. Andreas Plückthun (Vorsitz)

Prof. Dr. Donald Hilvert

Prof. Dr. Hans Rudolf Bosshard

Zürich 2005

Acknowledgement

Publications

Amstutz P., Forrer P., Zahnd C. & Pluckthun A. (2001) *In vitro* display technologies: novel developments and applications. Curr. Opin. Biotechnol. Aug, **12**, 400-5.

Schaffitzel C., Zahnd C., Amstutz P., Luginbühl B. & Plückthun A. (2001) *In vitro* Selection and Evolution of Protein-Ligand Interactions by Ribosome Display. (Golemis, E., ed) 535-567, Cold Spring Harbor Laboratory Press, New York

Amstutz P., Pelletier J.N., Guggisberg A., Jermutus L., Cesaro-Tadic S., Zahnd C. & Plückthun A. (2002) *In vitro* selection for catalytic activity with ribosome display. J. Am. Chem. Soc. **124**, 9396-403

Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003) Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489-503.

Zahnd C., Spinelli S., Luginbuhl B., Amstutz P., Cambillau C. & Plückthun A. (2004) Directed *in Vitro* Evolution and Crystallographic Analysis of a Peptide-binding Single Chain Antibody Fragment (scFv) with Low Picomolar Affinity. J. Biol. Chem. **279**, 18870-7.

Binz H.K.*, Amstutz P.*, Kohl A.*, Stumpp M.T., Briand C., Forrer P., Grütter M.G. & Plückthun A. (2004) High-affinity binders selected from designed ankyrin repeat protein libraries. Nat. Biotechnol. **22**, 575-82.
*These authors contributed equally to this work

Amstutz P., Binz H.K., Zahnd C. & Plückthun A. Ribosome Display: *In vitro* Selection for Protein-Protein Interactions. (Celis, J.E., ed.), CELL BIOLOGY: A Laboratory Handbook, in press

Amstutz P.*, Binz H.K.*, Parizek P., Stumpp M., Kohl A., Briand C., Grütter M.G., Forrer P. & Plückthun A. (2005) Intracellular Kinase Inhibitors selected from designed repeat protein libraries. J. Biol. Chem. **280**, 24715-221
*These authors contributed equally to this work

Kohl A.*, Amstutz P.*, Parizek P., Binz H.K., Capitani G., Forrer P., Plückthun A. & Grütter M.G. Allosteric Inhibition of Aminoglycoside Phosphotransferase by a Designed Ankyrin Repeat Protein. Structure **13**, 1131-41
*These authors contributed equally to this work

Amstutz P.*, Koch H.*, Binz H.K., Deuber S., Forrer P., Pavlovic J. & Plückthun A. Highly specific JNK2 binders selected by ribosome display and PCA from designed AR protein libraries. *To be Submitted*
*These authors contributed equally to this work

Abstract

Antibodies are indispensable binding molecules in research, diagnostics and therapy. Nevertheless, antibodies are afflicted with problems, such as poor expression and limited stability, especially in the intracellular milieu. Thus, superior alternatives would be of great value. This thesis describes the first successful selections of binding molecules and inhibitors from designed ankyrin repeat (AR) protein libraries.

In the first part of this thesis, designed AR proteins and ribosome display are introduced. Binding molecules based on AR proteins were devised to create an alternative to antibodies, overcoming some antibody-related problems. Designed AR proteins combine high affinity and specificity in target binding with high expression levels and high stabilities in all conditions. The selections of AR proteins binding to the given target proteins were achieved by ribosome display, an *in vitro* selection technology, ideally suited to identify or evolve binding molecules from combinatorial protein libraries. As no transformation step limits the size of the applicable library, this technology is ideal to handle very large libraries. Advances in the selection protocol were established, including an optimized approach for the generation of the library construction and the selection in the presence of detergents, allowing panning for membrane protein binders.

The second part of this thesis describes the initial work in selecting binding molecules and inhibitors from designed AR protein libraries. As general proof-of-concept, several different target proteins were chosen, including maltose binding protein (MBP) of *Escherichia coli* (*E. coli)*, aminoglycoside phosphotransferase (APH), the eukaryotic protein kinases p38, JNK1, JNK2 and AMPK, further, Sec YEG, a membrane protein of *E. coli* responsible for protein export and finally a non-structural protein of Semliki Forest Virus. Against these target proteins, specific binding molecules were obtained, demonstrating the power of designed AR protein libraries. The crystal structure of a selected AR protein in complex with its target protein (MBP) visualized the binding interaction at atomic resolution and revealed an interface indistinguishable from natural protein-protein interactions.

To demonstrate intracellular functionality of designed AR proteins, inhibitors against APH were selected. APH is a bacterial kinase and mediates kanamycin resistance in different pathogenic bacteria. A combined selection-screening approach yielded AR proteins that inhibited APH intracellularly, conferring

kanamycin sensitivity. The most potent of the selected inhibitors mediated a phenotype comparable to the APH gene knock out. The selected AR proteins also inhibited the enzyme *in vitro*, and the binding constants were in the low to sub-nanomolar range. Detailed biochemical characterization of the most potent inhibitor, including structure determination of the enzyme/inhibitor complex rationalized the inhibition mechanism. The AR protein bound and stabilized the kinase in an unproductive conformation, thus causing inhibition.

To monitor the ribosome-display selection process, to further enrich the binders, and to again demonstrate intracellular efficacy of the AR proteins, protein complementation assay (PCA) was applied after each ribosome-display selection round for JNK2 binding molecules. The obtained results indicated that only two rounds of ribosome display are sufficient to select nanomolar binders, and that an enrichment factor of 10^6 per round was achieved. All selected binders were fully functional in the cellular cytoplasm. The specificity of the selected AR proteins was very high, as they only interacted with the target protein JNK2, but did not bind to JNK1, which shows 81% identity on the protein sequence level.

In conclusion, this thesis demonstrates that designed AR proteins are a true alternative to antibodies, surpassing them in several biophysical properties. The selected AR proteins bound their targets with high specificity and affinity, were very stable, fully functional inside the cell and successful in crystallization trials.

These results open new possibilities in the fields of target validation, functional and structural genomics and might even be of use for diagnostic and therapeutic applications.

Zusammenfassung

Antikörper sind unersetzliche Bindemoleküle in der Forschung, der Diagnose und der Therapie. Trotzdem sind Antikörper mit verschiedenen Problemen behaftet, so wie teure Produktion und limitierte Stabilität, vor allem in der Zelle. Deshalb wären bessere Moleküle von grossem Wert. Diese Arbeit beschreibt die erste erfolgreiche Selektion von Bindemolekülen und Inhibitoren aus „Designed Ankyrin Repeat (AR) Protein" Bibliotheken.

Im ersten Teil dieser Arbeit werden „Designed AR Proteine" und „Ribosome Display" vorgestellt. AR Protein-basierte Binder waren konzipiert worden, um eine Alternative zu Antikörpern zu schaffen, welche die antikörperspezifischen Probleme beheben würde. „Designed AR Proteine" sollten hochaffines und hochspezifisches Binden, mit guter Expression und hoher Stabilität, auch unter reduzierenden Bedingungen, verbinden. Die Selektionen von Bindern aus AR Protein Bibliotheken wurden mittels „Ribosome Display" bewerkstelligt. Diese *in vitro* Selektionstechnologie ist ideal um aus Proteinbibliotheken Binder zu selektionieren. Da kein Transformationschritt die Biblothekgrösse begrenzt, ist „Ribosome Display" ideal um sehr grosse Bibliotheken in Selektionen zu verwenden. Aktuelle Verbesserungen im Selektionsprotokoll werden hier beschrieben, welche eine optimierte Strategie zur Bibliothekskonstruktion beinhalten, sowie die Möglichkeit zur Selektion in Gegenwart von Detergenzien, was die Selektion von Membranproteinbindern erlaubt.

Der zweite Teil dieser Arbeit beschreibt die ersten Erfolge bei der Selektion von Bindern und Inhibitoren aus den „Designed AR Protein" Bibliotheken. Eine Vielzahl von Zielmolekülen wurde gewählt, unter anderem Maltose-binde-Protein (MBP) von *Escherichia coli (E. coli)*, aminoglycoside phosphotransferase (APH), die eukaryotischen protein kinasen p38, JNK1, JNK2 und AMPK, weiter Sec YEG, ein Membranprotein aus *E. coli*, verantwortlich für den Proteinexport und schliesslich nicht-struktur-assoziierte Protein des Semliki Forest Virus. Gegen sämtliche dieser Zielmoleküle wurden spezifische Binder generiert, was die Qualität der AR Proteinbibliothek unterstreicht. Eine Kristallstruktur von einem AR Protein in Komplex mit seinem Zielmolekül visualisierte die Bindung mit atomarer Auflösung und beschtätigte eine natürliche Protein-Protein Interaktion.

Der nächste Schritt war die Demonstration der intrazellulären Funktionalität der AR Proteine. Zu diesem Zweck wurde APH als Zielmolekül ausgewählt. APH ist eine bakterielle Kinase, die pathogenen Bakterien Kanamycin Resistenz vermittelt. Ein kombiniertes Selektions-Screening Verfahren brachte

verschiedene AR Proteine hervor, die APH inhibierten und dabei die Kanamycin Sensitivität wiederherstellten. Die potentesten Inhibitoren erwirkten beim Bakterium einen Phänotyp, wie er bei Abwesenheit des APH Gens auftritt. Die selektionierten Inhibitoren inhibierten das Enzym auch *in vitro* und die Affinität war im tief- bis sub-nanomolaren Bereich. Der Inhibitionsmechanismus des besten Inhibitors wurde mittels einer detaillierten biochemischen und strukturellen Analyse aufgeklärt. Das AR Protein interagierte und stabilisierte eine nicht aktive Konformation des Enzyms

Um den Selektionsprozess zu erfassen, um die Binder weiter anzureichern und um nochmals die intrazelluläre Funktionalität der AR Proteine zu zeigen wurde ein Protein-Komplemetationstest nach jeder Ribosome-Display Selektionsrunde gegen JNK2 verwendet. Die Resultate zeigten, dass nur zwei Selektionsrunden nötig waren, um Binder mit nanomolarer Affinität zu selektionieren. Dies entspricht einem Anreicherungsfaktor von etwa 10^6 pro Selektionsrunde. Alle identifizierten JNK2 Binder waren funktional in der Zelle und waren sehr zielmolekülespezifisch; sie erlauben zwischen JNK2 und JNK1 zu unterscheiden, die zu 81% identisch sind.

Diese Resultate zeigen klar auf, dass es uns gelungen ist eine vielversprechende Alternative zu Antikörpern zu generieren, welche diese in wichtigen Punkten sogar übertrifft. Die selektionierten AR Proteine interagierten mit den entsprechenden Zielmolekülen mit hoher Affinität und Spezifität, waren sehr stabil, funktional in der Zelle und äusserst erfolgreich in Kristallisationsansätzen.

Zusammenfassend lässt sich sagen, dass die AR Protein Technologie neue Möglichkeiten in den Bereichen Target Validation, Genomprojekten und möglicherweise sogar in der Diagnose und der Therapie öffnet.

Contents in Brief

Contents

Introduction

Contents

General Remarks

Proteins are fundamental to life. They play crucial roles in nearly all biological processes: enzymes catalyse chemical reactions, structural proteins are responsible for cellular architecture, receptors are involved in signal transduction and antibodies fight intruders. Understanding life at a molecular level will involve the understanding of proteins. This thesis addresses proteins at different levels.

First, technologies are discussed, which allow the engineering of proteins. The goal is to understand proteins and to improve them, such that they might be applied to biological problems. Two approaches dominate the field of protein engineering: "rational design" and "directed evolution". Both approaches have been applied successfully and might be combined, as they are truly complementary.

Second, the discussed methods of protein engineering are applied to generate binding molecules. Up to date, the binding molecule of choice was the antibody. As antibodies were almost the only binding molecules, which could specifically bind any given target with high affinity and specificity, they were used in a wide range of applications, including research, diagnostics and therapy. Nevertheless antibodies suffer from limitations originating from their biophysical properties. A valid alternative is presented: designed Ankyrin repeat (AR) proteins. These proteins bind their targets as antibodies do, but overcome the limitations of antibodies.

The generation of AR protein as antibody alternative was achieved by a combination of rational design and directed evolution, giving evidence for the power of the synergy of these approaches.

1. Protein Engineering

The distinct amino acid sequence of a protein defines its three dimensional structure and with that its function. Therefore, mutations in the sequence can alter structure and function. The goal of protein engineering is to change the

sequence of a given protein to improve its biophysical characteristics. Different approaches can be chosen to achieve this goal: "rational design", "directed evolution" or a combination thereof.

1.1 Rational design

In the rational design approaches, information about a protein (or a class thereof) is collected and rationalized to predict beneficial mutations. The more information about the target protein (and proteins in general) is available, the higher the probability of success. These *in silico* methods can be applied to the generation of novel binding specificities (Looger et al., 2003), enzyme engineering (Dwyer et al., 2004), addressing protein structure and folding problems (Kuhlman et al., 2003, Ventura & Serrano, 2004) and engineering of protein stability (Steipe et al., 1994, Kuhlman & Baker, 2004).

Recently, different groups published remarkable results. Hellinga and co-workers used rational design to generate specific high-affinity receptors, based on periplasmic binding protein (PBP). This protein superfamily is known to bind a wide variety of different ligands and many high resolution structures with ligands bound have been determined (Dwyer et al., 2004). The redesign to alter the ligand specificity of the PBP comprised the following steps: First, all side chains interacting with the natural ligand were identified and replaced by alanine, defining the so called "primary coordination sphere" (PCS). An ensemble of possible conformations of a putative ligand was then placed in the potential binding pocket and its most promising conformations were fixed. Using a side chain rotamer library and a dead-end elimination algorithm, a PCS for each ligand conformation was calculated. Optimal sequences were identified determining the global energy minimum and ranked according to dead-end elimination energy, protein-ligand van der Waals contact energy, number of unsatisfied hydrogen bonds and ligand exposed surface area. Finally, a small number of top-ranked designs were tested experimentally. By this approach specific high affinity receptors to different ligands were successfully generated, which were fully functional *in vivo* (Looger et al., 2003). The same group even went one step further and designed enzymatic activity into a PBP (Dwyer et al., 2004). A three step design approach was applied to generate a novel enzyme.

First, key interactions for catalysis, which were geometrically fixed, were defined. Second, positions for the catalytic side chain were identified by a combinatorial approach. Third, the remaining complementary surface was generated around the defined substrate. This approach yielded PCB variants that showed triose phosphate isomerase activity with a 10^5-10^6 fold rate enhancement compared to the uncatalyzed reaction. The most active PCB variant was able to complement the wild-type enzyme in *E. coli*, supporting growth under glugoneogenic conditions.

Another impressive example of the power of rational design was the prediction of an artificial protein, which had all the hallmarks of a natural one. Baker and co-workers applied an approach that iterates between sequence optimisation and structure prediction. A model of a novel protein topology, which was not found in the Protein Data Bank (PDB), was defined. Three dimensional starting models were generated by assembling three- and nine-residue fragments from the PDB with secondary structures consistent with defined novel protein topology. Then, the sequence and the structure were simultaneously optimized by cycling between sequence design and structure optimization. The crystal structure of the designed protein matched the predicted one almost perfectly (Kuhlman et al., 2003).

While the energy minimizing approaches are based on an understanding of the three-dimensional structure, another approach, termed consensus design, makes use of sequence information. Here, the alignment of homologous sequences is used to define a consensus sequence with improved characteristics (Steipe et al., 1994, Forrer et al., 2004). The underlying idea is that the functionally important amino acids are more conserved than others. Comparing all available sequences will define the most conserved choice for each position, resulting in the optimal sequence. This approach is very dependent on the sequences included in the alignment process, with chances for success rising with the number of available sequences. As more and more sequence data are available from genome sequencing projects, this approach gains strength.

1.2 Directed Evolution

Natural evolution describes a simple process of optimization. Every organism carries its master plan as genes encoded on DNA (genotype). These genes encode proteins, which are the actual working tools (phenotype). Mutations occur spontaneously and at random positions at the DNA level and can manifest themselves as altered proteins. If, by chance, a mutation is beneficial for the organism, this individual is selected by the criterion of fitness and will pass its improved master plan on to the next generation.

This natural evolutionary principle of succeeding rounds of diversification followed by selection of the fittest can be mimicked in the test tube to evolve proteins with improved characteristics in an approach called "directed evolution" (Plückthun et al., 2000; Farinas et al., 2001). In an experimental set-up, the genotype must be linked to the phenotype, as seen in nature, and a selection pressure must be applied to a starting population, to identify potential candidates. Short generation times combined with large and diverse starting populations allow more individuals to enter the game and therefore raise the chances of success.

2. Binding Molecules

As more and more genome sequencing projects are completed, the number of identified genes is rising drastically (Venter et al., 2004). The main task now lies in the assignment of function to the proteins encoded in these genes. Specific binders to these proteins would facilitate their investigation, allowing determination of localization *in vivo*, detection on Western blots, immunoprecipitation experiments and possibly help structure determination of the target protein. In some cases the binding molecule might even find use in diagnostic applications or serve as a lead compound in the drug discovery process. Specific binding molecules are often proteins, in most cases antibodies.

2.1 Antibodies

The most common binding proteins are antibodies. These proteins are naturally produced by B-lymphocytes as key elements of the humoral immune response of all vertebrates to fight intruders. Thereby antibodies have to be able to discriminate between self and non-self. Nature achieves this by creating antibody libraries, which are depleted of "self"-recognizing antibodies. During B-lymphocyte maturation antibody gene rearrangement creates stretches coding for variable surface loops on the antibody framework, which are termed complementary determining regions (CDRs) (Tonegawa, 1983). The CDRs of an antibody can interact with the target and are therefore responsible for specific binding. Every single B-lymphocyte expresses a distinct antibody. During B-cell maturation "self" recognizing individuals are negatively selected, all others have the potential to recognize an intruder. The human body produces around 10^9 distinct B-lymphocytes (Winter, 1998). This large diversity of naturally produced antibodies can be used to select target-specific antibodies by immunization. The target protein is injected into a laboratory animal, which induces an immune response and leads to production of specific antibodies. Nevertheless antibodies suffer from limitations: They are built from several domains and suffer from high production costs. Furthermore, they contain stabilizing disulfide bonds, which can not form in the reducing environment of the cellular cytoplasm, making intracellular applications very difficult (Biocca & Cattaneo, 1995).

2.2 Alternative scaffolds - designed ankyrin repeat proteins

The natural principle of generating a pool or a library of proteins with a randomized surface area can be transferred from antibodies to other proteins (reviewed in Skerra, 2000). An ideal alternative to antibodies would be built of one small domain, would be well expressed and contain no cysteines, while still retaining the beneficial properties of antibodies such as tight and specific binding. The approach described here is based on the ankyrin repeat (AR) protein as scaffold. AR proteins are, next to antibodies, the most abundant class of binding molecules found in nature (Sedgwick & Smerdon, 1999). Unlike

antibodies they do not contain any disulfide bonds and are found in all cellular compartments. AR proteins occur in all phyla and mediate protein-protein interactions with affinities in the nanomolar to picomolar range (Binz et al., 2003). They often act as natural intracellular modulators of many biological processes (Brotherton et al., 1998; Russo et al., 1998). AR proteins are built of consecutive homologous repeats, which assemble to form elongated proteins with a continuous hydrophobic core.

Our "rational design" approach combined structure and consensus analyses to engineer an AR module of 33 amino acids with fixed framework residues and randomized potential interaction residues. This module was used as a building block to create combinatorial libraries of designed AR proteins of different sizes, with diversities greater than 10^{10} (Binz et al., 2003). The same approach was also applied to Leucine rich repeat proteins (Stumpp et al., 2003). The unique modular architecture of repeat proteins allows module shuffling, insertion and deletion. Furthermore, the binding surface is not restricted in its dimension, as it may be adapted to the target by addition or removal of repeat modules (Forrer et al., 2003).

In all approaches based on antibodies or alternative frameworks, a pool of proteins, termed a protein library, is used for binder generation. Usually, the diversified amino acids of a protein library are not distributed randomly over the whole protein, but are located in surface loops or other well-defined locations (Figure 1). The constant part of the protein is called the framework, which ensures the correct protein fold, while the randomized positions are potentially involved in binding. In this way, all individuals of the library have the same overall topology, showing the protein class specific fold, but have a randomized surface for potential target interaction (Figure 1).

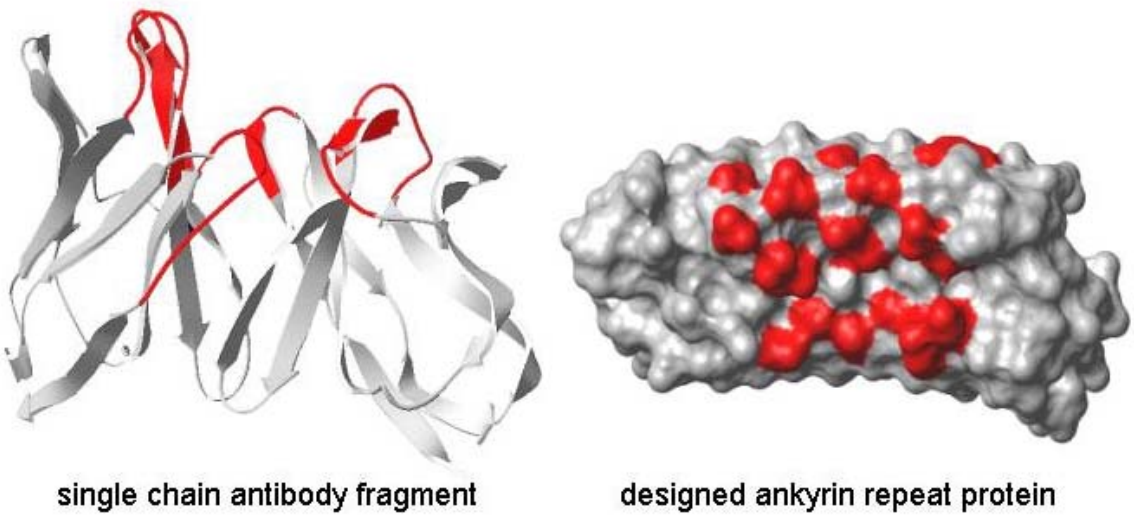


Figure 1. Comparison of a single chain antibody fragment (scFv) and a designed ankyrin repeat (AR) protein. The scFv is depicted in ribbon representation, while the AR protein is depicted in space-fill representation. For both the conserved framework residues are colored in grey, while the randomised potential interaction residues are in red.

3. Selecting Binding Molecules

Thirty years ago, immunization of test animals was the only way to generate target specific antibodies (Köhler & Milstein, 1975). This procedure was not optimal. Some targets were not immunogenic, others were not stable in the laboratory animal or simply too toxic, failing to give rise to specific binders. In any case, the selection process, taking place in the animal, was a black box, virtually uncontrollable, with an uncertain outcome. To overcome these limitations and to put the selection process in the hands of the researcher, novel selection methods were developed. These methods also opened the door for selection of proteins other than antibodies.

The starting point of every selection is a protein library. These libraries are used for selections to pick out those proteins which show specific binding. In analogy to the natural evolution principle, the library is the starting population. The larger and more diverse the starting population is, the higher the probability will be to find an individual with the desired properties.

3.1 *In vivo* selection technologies

In vivo selection technologies link the desired binding event to cellular survival. A cellular protein, with a selectable phenotype, is genetically divided in two parts and the potential interaction partners are genetically fused to each part. If they indeed interact and form a complex, the divided protein is reunited, is active and gives rise to a selectable phenotype. This principle is applied in the protein complementation assay (PCA), yeast two hybrid, bacterial two-hybrid and many other systems (Fields & Song, 1989; Michnick et al., 2000). In any case the interaction occurs *in vivo* and the precise selection conditions are difficult to control.

3.2 *In vitro* selection technologies - Phage Display

To gain control over the selection conditions, the selection step must happen *in vitro*. For this purpose *in vitro* selection technologies were established. The most prominent one is phage display (Smith, 1985). The protein of interest is genetically fused to a phage coat protein. Every phage displays a distinct protein, carrying the corresponding gene inside, thereby achieving the necessary coupling of genotype and phenotype. The whole population of phages, each displaying a different protein variant, can be incubated with immobilized target protein and the phages displaying a protein interacting with the target can be captured. While the selection process now is carried out *in vitro*, phage production still happens *in vivo*. This brings limitations in respect to the size of the starting population due to transformation limitations.

3.3 Complete *in vitro* selection technologies - Ribosome Display

Complete *in vitro* display technologies overcome these limitations. These systems are based on *in vitro* translation to achieve the necessary coupling of genotype and phenotype by different experimental tricks. In this thesis the focus lies on ribosome display (Hanes & Plückthun, 1997; He & Taussig, 1997; Mattheakis et al., 1994). Ribosome display relies on non-covalent ternary

complexes, which are formed during *in vitro* translation. They contain mRNA (genotype), ribosome and nascent polypeptide (phenotype). The coding sequence of the investigated protein is genetically fused to a C-terminal tether, such that the domain of interest can fold while the tether is still in the ribosomal tunnel. This fusion construct lacks a stop codon at the mRNA level, thus preventing release of the mRNA and the polypeptide from the ribosome. High concentrations of magnesium and low temperatures further stabilize the ribosomal complex. These complexes can be used directly to select for properties of the displayed protein. This complete *in vitro* technology allows the handling of very large starting populations and shortens the selection time to a minimum.

4. Combining Approaches

As mentioned above the approaches used for of protein engineering can be divided into two categories: "rational design" and "directed evolution". The debate about which approach is superior is ongoing. In principle this discussion is out of place, as "rational design" and "directed evolution" are absolutely complementary to each other (van Regenmortel, 2000). One central drawback of rational design is that the outcome sequences, defined *in silico*, will still have to be tested in a wet lab, limiting throughput. Here, "directed evolution" methods can be applied to test more sequences, raising the chances for success. The outcome of "directed evolution" experiments, on the other hand, is directly linked to the sequences in the starting population. If there are no good individuals present in the first place, they cannot be selected. Here "rational design" can improve the quality of the starting population, increasing the chances of finding the desired clones.

The work presented here is one example of the synergistic combination of both approaches for the successful generation of novel binding proteins. Rational design was applied to generate combinatorial AR protein libraries, ensuring a high quality population of potential binders (Binz et al., 2003). These libraries were the starting point for selection experiments by ribosome display, allowing the rapid selection of binders from very large libraries.

Proof-of-concept selection experiments yielded AR protein binders to a whole range of different target proteins. The binding mode was analyzed with respect to specificity and affinity, showing specific binding in the nanomolar range. The crystal structure of a complex, visualizing the binding interface at atomic resolution, revealed a protein-protein interaction, indistinguishable from natural ones.

The potential of intracellular applications was demonstrated by selection of AR proteins inhibiting a bacterial kinase responsible for kanamycin resistance. The selected inhibitors created an *in vivo* "knock down" showing kanamycin sensitivity. The most potent enzyme inhibitors were almost as efficient as the total gene knock out. Again, a crystal structure helped to rationalize the binding and inhibition mechanism.

Finally, the ribosome-display selection process for JNK2 binding molecules was monitored by a protein complementation assay. The obtained results indicated that only two rounds of ribosome display are sufficient to select nanomolar binders, and that an enrichment factor of 10^6 per round was achieved. All selected binders were fully functional in the cellular cytoplasm and the specificity of the selected AR proteins was very high as they only interacted with the target JNK2, but did not bind to JNK1, which shares 81% sequence identity on the protein sequence level.

Our results suggest that AR proteins are a valid alternative to antibodies, with equal or superior characteristics. The described approach demonstrates the power of the combination of "rational design" and "directed evolution". Having completed the proof-of-concept experiments, the focus is now on generating high quality binding molecules for research, diagnostics and therapeutics.

5. References

Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J. Mol. Biol. 332, 489-503.

Biocca, S. & Cattaneo, A. (1995). Intracellular immunization: antibody targeting to subcellular compartments. Trends Cell. Biol. 5, 248-252.

Brotherton, D. H., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P. J., Volyanik, E., Xu, X., Parisini, E., Smith, B. O., Archer, S. J., Serrano, M., Brenner, S. L., Blundell, T. L. & Laue, E. D. (1998). Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. Nature 395, 244-250.

Dwyer, M. A., Looger, L. L. & Hellinga, H. W. (2004). Computational design of a biologically active enzyme. Science 304, 1967-1971.

Farinas, E. T., Bulter, T. & Arnold, F. H. (2001). Directed enzyme evolution. Curr. Opin. Biotechnol. 12, 545-551.

Fields, S. & Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature 340, 245-246.

Forrer, P., Binz, H. K., Stumpp, M. T. & Plückthun, A. (2004). Consensus design of repeat proteins. ChemBioChem 5, 183-189.

Forrer, P., Stumpp, M. T., Binz, H. K. & Plückthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett. 539, 2-6.

Hanes, J. & Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. U S A 94, 4937-4942.

He, M. & Taussig, M. J. (1997). Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites. Nucleic Acids Res. 25, 5132-5134.

Köhler, G. & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256, 495-497.

Kuhlman, B. & Baker, D. (2004). Exploring folding free energy landscapes using computational protein design. Curr. Opin. Struct. Biol. 14, 89-95.

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science 302, 1364-1368.

Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. Nature 423, 185-190.

Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. (1994). An in vitro polysome display system for identifying ligands from very large peptide libraries. Proc. Natl. Acad. Sci. U S A 91, 9022-9026.

Michnick, S. W., Remy, I., Campbell-Valois, F. X., Vallee-Belisle, A. & Pelletier, J. N. (2000). Detection of protein-protein interactions by protein fragment complementation strategies. Methods Enzymol. 328, 208-230.

Plückthun, A., Schaffitzel, C., Hanes, J. & Jermutus, L. (2000). In vitro selection and evolution of proteins. Adv. Protein Chem. 55, 367-403.

Russo, A. A., Tong, L., Lee, J. O., Jeffrey, P. D. & Pavletich, N. P. (1998). Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. Nature 395, 237-243.

Sedgwick, S. G. & Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem. Sci. 24, 311-316.

Skerra, A. (2000). Engineered protein scaffolds for molecular recognition. J. Mol. Recognit. 13, 167-187.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228, 1315-1317.

Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. J. Mol. Biol. 240, 188-192.

Stumpp, M. T., Forrer, P., Binz, H. K. & Plückthun, A. (2003). Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. J. Mol. Biol. 332, 471-487.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature 302, 575-581.

van Regenmortel, M. H. (2000). Are there two distinct research strategies for developing biologically active molecules: rational design and empirical selection? J. Mol. Recognit. 13, 1-4.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. & Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66-74.

Ventura, S. & Serrano, L. (2004). Designing proteins from the inside out. Proteins 56, 1-10.

Winter, G. (1998). Synthetic human antibodies and a strategy for protein engineering. FEBS Lett. 430, 92-94.

Chapter 2

Designed Ankyrin Repeat Protein Libraries

Contents

doi:10.1016/S0022-2836(03)00896-9 *J. Mol. Biol.* (2003) **332**, 489–503

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins

H. Kaspar Binz, Michael T. Stumpp, Patrik Forrer, Patrick Amstutz and Andreas Plückthun*

Biochemisches Institut
Universität Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland

We describe an efficient way to generate combinatorial libraries of stable, soluble and well-expressed ankyrin repeat (AR) proteins. Using a combination of sequence and structure consensus analyses, we designed a 33 amino acid residue AR module with seven randomized positions having a theoretical diversity of 7.2×10^7. Different numbers of this module were cloned between N and C-terminal capping repeats, i.e. ARs designed to shield the hydrophobic core of stacked AR modules. In this manner, combinatorial libraries of designed AR proteins consisting of four to six repeats were generated, thereby potentiating the theoretical diversity. All randomly chosen library members were expressed in soluble form in the cytoplasm of *Escherichia coli* in amounts up to 200 mg per 1 l of shake-flask culture. Virtually pure proteins were obtained in a single purification step. The designed AR proteins are monomeric and display CD spectra identical with those of natural AR proteins. At the same time, our AR proteins are highly thermostable, with T_m values ranging from 66 °C to well above 85 °C. Thus, our combinatorial library members possess the properties required for biotechnological applications. Moreover, the favorable biophysical properties and the modularity of the AR fold may account, partly, for the abundance of natural AR proteins.

Keywords: ankyrin repeat; combinatorial library; consensus sequence; protein design; scaffold

*Corresponding author

Introduction

Repeat proteins mediate numerous key protein–protein interactions in nature.[1,2] Their repetitive architecture permits the adaptation of their size and thus their variable and modular binding surface to a target protein, leading to high-affinity interactions. We developed a strategy that exploits this modular architecture for the generation of combinatorial libraries of repeat proteins with novel binding specificities (Figure 1).[3] Our strategy consists of designing a self-compatible repeat module for a given repeat type. In such a repeat module, residues important for maintaining the repeat structure (i.e. framework residues) are defined, while potential target interaction residues are randomized. Such self-compatible repeat modules can then be joined repetitively to yield a stack of repeats. To form repeat domains, the continuous hydrophobic core of this stack is sealed by N and C-terminal capping repeats (Figure 1). We hypothesized that, by using this strategy, libraries of repeat proteins of different lengths with very large and highly diversified interaction surfaces could be generated. Our strategy is thus in clear contrast to the traditional scaffold approach, which would consist of randomizing the surface or loops of a well characterized natural protein whose dimensions are fixed.[4,5]

Ankyrin repeat (AR) proteins constitute a very attractive class of repeat proteins to test our strategy. AR proteins mediate many important protein–protein interactions in virtually all species,[6] and are found intracellularly, extracellularly

Abbreviations used: AR, ankyrin repeat; IMAC, immobilized metal-ion affinity chromatography; PDB, Protein Data Bank.

E-mail address of the corresponding author: plueckthun@bioc.unizh.ch

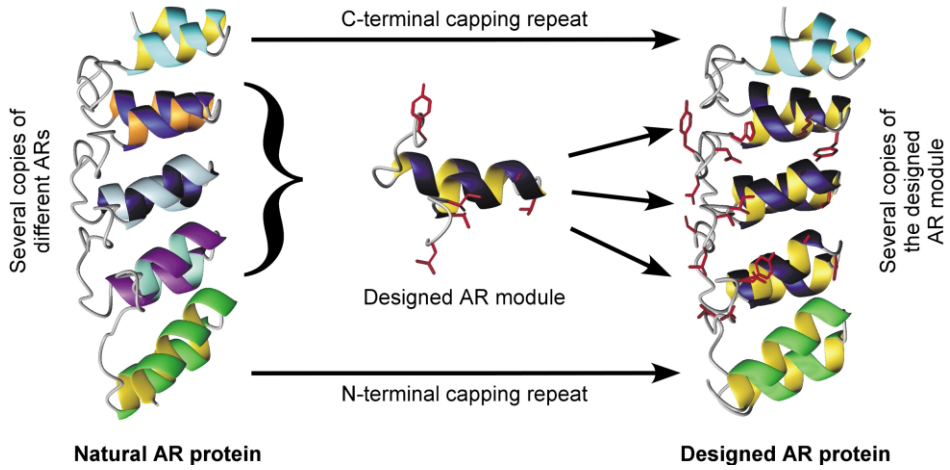


Figure 1. The strategy to generate designed AR protein libraries. From structure and sequence alignments of natural ARs, a self-compatible AR module is designed. This repeat module consists of fixed framework residues and randomized potential interaction residues (shown in red stick mode). Various numbers of this AR module (here three) are then cloned between N and C-terminal capping repeats. By using this strategy, combinatorial libraries of designed AR proteins of varying repeat numbers can be generated. The randomized positions on several adjacent repeats create a large potential interaction surface presented on a rigid AR scaffold. This Figure was created using PDB entries 1AWC[21] and 1MJ0[29] with the help of MOLMOL.[54]

and in membrane-bound form, indicating that these proteins can adapt to many different environments. The fact that there are more than 2000 known AR proteins (>14,000 ARs)[7] underscores their importance in nature. AR domains are composed of stacked ARs, consisting typically of 33 amino acid residues, each forming a β-turn followed by two antiparallel helices and a loop reaching the β-turn of the next repeat (Figure 2).[8,9] Usually, four to six repeats[6] assemble into domains, but the crystal structure of ankyrin R, consisting of 12 ARs in a single domain, was reported recently,[10] indicating that there is virtually no limit to the number of repeats that can fold in one AR domain.

Even though different parts of the surface of AR domains could, in principle, be involved in protein–protein interactions,[10] most AR domains interact with their cognate partners *via* the protruding β-turns and the following α-helices. Typically, several adjacent repeats establish contact. This patch-wise interaction mode leads to high-affinity interactions, exemplified by the mouse GA-binding protein (GABP) β1 binding GABPα with a K_D of 0.78 nM or by IκBα inhibiting the DNA binding of NF-κB with a K_i of 3.1 nM.[11,12]

We implemented our novel strategy on AR proteins and generated combinatorial libraries of designed AR proteins of distinct repeat numbers. Here, we describe the design, construction and analysis of these libraries. The accompanying paper by Stumpp *et al.*[13] describes the implementation of our strategy on leucine-rich repeat proteins, another abundant repeat protein class.

Results

A key step in our strategy (Figure 1) is the design of self-compatible repeat modules.[3] This should ensure the proper stacking of the repeat modules into repeat domains. ARs feature a high degree of sequence and structure conservation and, thus, one way to generate self-compatible repeats is through consensus design. In consensus design, the conserved intra- and interrepeat interactions characteristic for the AR domain fold are implemented into the repeat module. In addition, consensus design may lead to improved repeat stability.[3] We describe here the design of a consensus AR, in which conserved framework residues are fixed and in which potential target interaction residues are randomized. The design is based on sequence and structure consensus analyses.

AR consensus sequence definition using sequence databases

The first important task for our consensus design was to choose an appropriate sequence data set. The SMART database[7] provided a large number of functionally unbiased AR sequences and was therefore chosen as the starting point. Numbering of the positions of the AR consensus was adapted from that used by Sedgwick & Smerdon.[9] The Clustal W[14] alignment of ARs was downloaded from SMART. The data set was further reduced to those sequences matching the length of the 33 amino acid residue consensus described earlier.[6,9,15,16] Only repeats without extra insertions or deletions were considered for the consensus definition. The resulting alignment of 229 ARs yielded consensus A containing residues 3–32 (Figure 3). To further refine consensus A and to define the lacking residues (1, 2 and 33), consensus A was circularly permutated and the lacking residues and those without clear preference (frequency

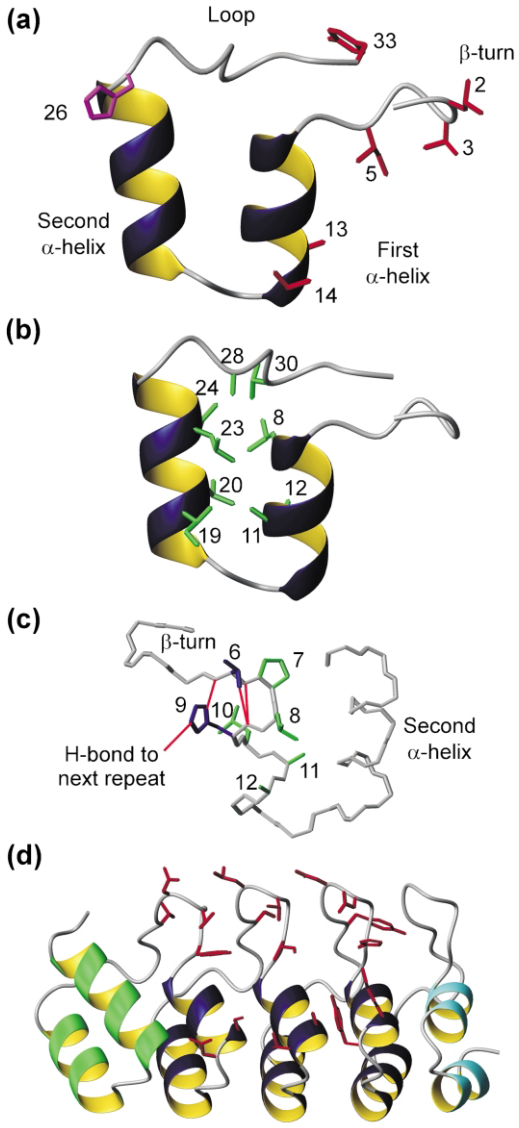


Figure 2. Crystal structure of the unselected N3C library member E3_5 (PDB entry: 1MJ0)[29] illustrated with MOLMOL.[54] (a) Potential interaction residues of the middle AR module (residues 77–109) are displayed in red on the AR framework in ribbon representation. The potential interaction residues are located in the β-turn and the concave surface of the L-shaped repeat. The partly randomized framework position 26 is displayed in magenta. The structural elements of the AR are labeled. (b) Hydrophobic framework residues and alanine residues pointing into the core of the middle AR module are colored in green on the AR in ribbon representation. (c) A rotated view of this middle AR module, which shows more clearly the TPLHLAA motif (residues 6–12) of the first α-helix with its characteristic H-bond pattern. Hydrophobic residues and alanine residues are colored in green, Thr6 and His9 are colored in blue and H-bonds are colored in red. The loop has been cut in this representation. (d) Crystal structure of E3_5 displaying a large potential interaction surface built by the randomized positions (shown in red stick mode). The N and C-terminal capping repeats and the internal repeat modules are colored in green, light blue and dark blue, respectively.

≤30%) were chosen from the alignment of repeats of AR proteins with known structure. The resulting consensus B (Figure 3) was subjected to a BLAST[17] search against GenBank.[18] The resulting top 200 BLAST hits were manually aligned and analyzed, yielding consensus C (Figure 3). An alignment of 2220 AR sequences stored in the PFAM database[19] confirmed the choice of the most frequent amino acids of consensus C (threshold 30%).

AR consensus refinement using structural data

To extend our sequence database analyses, we decided to include structural data for the final refinement leading to consensus D (Figure 3). The structural analysis included the ten AR protein 3D structures 1YCS,[8] 1AP7,[20] 1AWC,[21] 1A5E,[22] 1IKN,[23] 1NFI,[24] 1MYO,[25] 1IHB,[26] 1DCQ[27] and 1SW6.[28] In a first step, the PDB files were used to define potential target interaction residues and framework residues.

Potential target interaction positions

From 3D structures of complexes of AR domains with target proteins (1AWC, 1YCS, 1IKN, 1NFI), target interaction residues were identified using NACCESS† by analyzing the change in solvent-accessible surface area of AR domain residues upon complex formation. Interactions mostly involve the β-turns and the first α-helices of the AR proteins, i.e. positions 2, 3, 5, 13, 14 and 33 of the repeats. In consensus D (Figure 3), these positions were permitted to contain any amino acid except glycine, proline (both structurally unfavorable) or cysteine (may form unwanted disulfide bonds). All other positions in the consensus were defined as framework residues, and we thus tried to assign defined amino acids to these positions.

Framework positions

Positions 1 and 4 were defined as Asp and Gly, respectively, as these residues are frequent (Asp1 37%, Gly4 75% in BLAST search) and form a network of H-bonds extending over consecutive β-turns.[29] Furthermore, position 4 asks for a positive ϕ angle, and is thus best accommodated by Gly. The motif TPLHL for positions 6 to 10 is highly conserved in natural ARs. Thr6 forms several H-bonds to His9 (Figure 2), Pro7 breaks into α-helix 1 and is in the hydrophobic core. Leu8 lies in the hydrophobic core, pointing towards helix 2 and towards the next repeat. In addition to making H-bonds to Thr6, His9 establishes H-bond contact to Ala32 and to the next repeat (randomized position 5; Figure 2). Leu10 points towards helix 1 of the previous repeat, and probably stabilizes the interface between two repeat modules. Since Leu10 is the least conserved

† http://wolf.bms.umist.ac.uk/naccess/

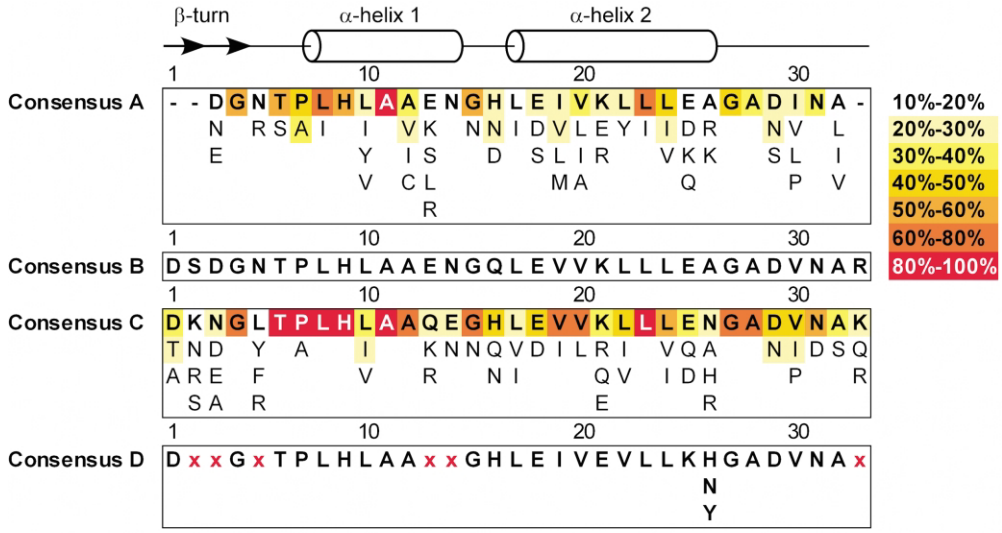


Figure 3. Stepwise definition of the AR consensus used in the present study. The amino acid frequency color code is indicated in the panel. For orientation, the secondary structure elements are indicated above the sequences. Consensus A was derived from an alignment of 229 ARs of the SMART database. It contained only residues 3–32 of the AR consensus. The sequence of consensus B was derived from consensus A, where lacking or non-conserved (cut-off ≤30%) residues were substituted by residues resulting from an alignment of repeats of AR proteins with known structure. A circularly permutated form of consensus B (starting from residue V20) was submitted for a BLAST search against GenBank. Through the circular permutation, residues 1, 2 and 33 could be analyzed and the consensus could be refined. Consensus C was derived from the BLAST search with consensus B. Structure-based considerations (see Results) led from consensus C to consensus D, the final sequence of the designed AR module. In consensus D, the potential target interaction residues are highlighted in red.

residue in the TPLHL motif (30% frequency in the BLAST search), and since it is, in part, surface exposed, it could, in principle, have been considered as potential target interaction residue. The TPLHL motif represents the N-terminal helix-cap of the first α-helix. The small hydrophobic helix-formers Ala11 and Ala12 are important for the overall shape of ARs (Figure 2). Their small size allows the repeat to be conical in form, narrow at the bottom and wide at the top.[10] Gly15 breaks out of helix 2 and has a positive ϕ angle. His16 is semi-buried and forms side-chain H-bonds to helix 1 of the previous repeat as well as backbone H-bonds to Ala11, Ile19 and Val20. Position 17 was not clear from database statistics (Leu being most prevalent with 26% in the BLAST search). From structural considerations, however, it is likely that a leucine residue would stabilize the repeat interface and could be the initiator of helix 2. For these reasons, and because of its high α-helical propensity,[30] we chose Leu at this position. Helix 2 is amphipathic and contributes to the overall shape of the AR. Larger amino acid residues in the hydrophobic core are situated closer to the loop, leading to increasing helix-helix distances. Several positions in helix 2 were not well defined in the consensus from database analyses and were thus assigned using other decision parameters. Glu18 was chosen because it occurs repeatedly in the cdk4 inhibitor p18 (chain B of PDB entry 1IHB)[26] and can thus apparently be tolerated multiple times. Similarly, Glu21 occurs repeatedly in GABPβ1 (chain B of PDB entry 1AWC, here called 1AWC_B).[21] Since positions 18 and 21 are well separated in space, both negative charges should be tolerated. Ile19 was chosen because it fits similarly as Val, but has a higher α-helical propensity.[30] Ile19 is part of the hydrophobic core (Figure 2), and so is Val20, which was defined from sequence analysis. Leu22, Leu23 and Leu24 constitute a rather conserved patch in the upper part of the hydrophobic core (Figure 2). However, position 22 was chosen to be Val, as this occurs repeatedly in GABPβ1. Lys25 was chosen, because it has the opposite charge of Glu21 and, as the latter, occurs repeatedly in GABPβ1. Arg could have also been chosen in this position. Position 26 was ambiguous, although Asn was most abundant. Ala was prominent in the alignment; however, to control the distance of the repeats, the amino acid at this position should fill enough space and should probably be polar. His was an alternative but there was the danger of creating a charge belt Lys25/His26 across the repeat domain. Tyr was another alternative suggested by GABPβ1. Finally, a combination of His, Tyr and Asn was chosen, since these amino acids can be encoded by the HAC codon. Gly27 breaks out of helix 2 having a positive ϕ angle and initiates the loop. Ala28 points into the hydrophobic core to anchor the loop (Figure 2), as do Val30 and Ala32. Asp29 and Asn31 are important for H-bond networking and keeping contact between the consecutive repeats.[29]

The final consensus D is displayed in Figure 3.

When checking pairs, triplets and quadruplets of amino acids of consensus D, all combinations occurred at least once in natural ARs, except the two quadruplets in the HLEIV sequence motif. Taken as a whole, consensus D was designed to encode self-compatible repeat modules that are built from conserved framework residues and randomized potential target interaction residues.

Design of capping repeats

Like other repeat proteins (see the accompanying paper[13]), natural AR domains have specialized terminal repeats (capping repeats) that function to seal the hydrophobic core of a stack of ARs (Figure 4). While the internal ARs have two hydrophobic repeat–repeat interfaces, the capping repeats have only one such interface and the exposed surface is hydrophilic (Figure 4). We thus reasoned that these capping repeats are needed in order to form a stable, well-folded AR domain, and we included capping repeats in our strategy.[3]

We decided to adapt naturally occurring capping ARs to our designed modules. The choice of appropriate capping repeats was based on two criteria: (i) the structure had to be known; and (ii) they had to be as compatible and thus as homologous as possible to our designed repeat module. When joining four, five or six designed repeat modules *in silico* and subjecting these sequences to BLAST searches against the PDB, mouse GABPβ1 was always the best hit. We thus decided to adapt the capping repeats of GABPβ1 for our design. While residues 1–26 of the N-terminal capping repeat, which form the two anti-parallel α-helices, were taken directly from 1AWC_B, the loop sequence GAPFT was changed to GADVNA. There were two reasons for this change: (i) modeling suggested that the bulky GAPFT loop did not sterically match the consensus GADVNA loop of the neighboring repeat. Phe in the GAPFT loop serves as a spacer between the N-terminal repeat and the second repeat of GABPβ1 and appears in combination with Ala in position 26 of the second repeat, where our design had adopted bigger residues. (ii) For cloning purposes, the end of the loop had to contain the sequence Asp-Val. The finally chosen sequence for the N-terminal capping repeat was therefore: DLGKKLLE AARAGQDDEV RILMAN GADV.

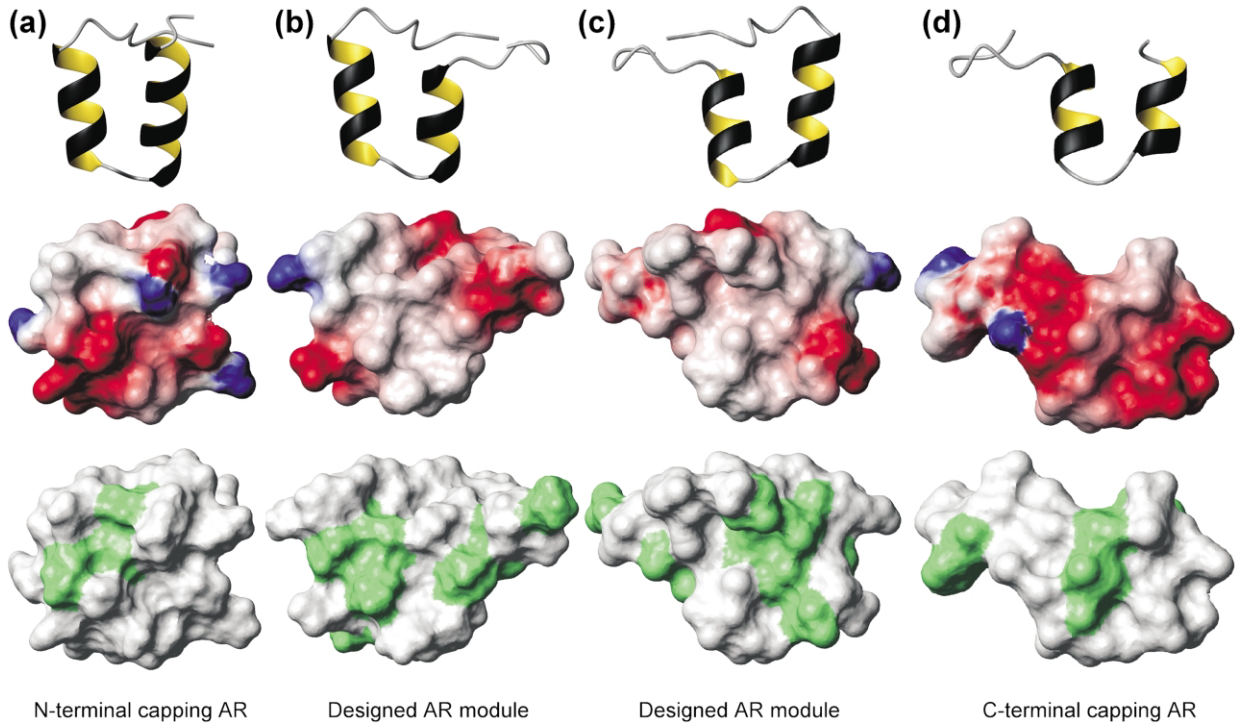


Figure 4. Charge distribution and hydrophobicity of AR surfaces. The charge distribution and hydrophobicity of the middle AR module of E3_5 (PDB entry: 1MJ0),[29] an unselected N3C library member, is compared to the N and C-terminal capping repeats of the same molecule. (a) The N-terminal capping repeat, seen in a lateral view from the N terminus, i.e. from the "outside" of the protein. (b) The middle repeat, seen from the same direction as in (a), i.e. exposing an otherwise buried surface. (c) The middle repeat, seen in a lateral view from the C terminus, i.e. an otherwise buried surface is shown. (d) The C-terminal capping AR seen in the same view as (c), i.e. from the "outside" of the protein. For orientation, the ARs are represented in ribbons on top. In the middle, charge representations are given with negative charges in red and positive charges in blue. The hydrophobicity is illustrated in the bottom row with hydrophobic side-chains in green. The solvent-exposed capping repeats have surfaces that are more charged than the repeat–repeat interfaces of the middle repeat. Likewise, larger hydrophobic areas characterize the repeat–repeat interfaces of the middle repeat in (b) and (c) compared to smaller hydrophobic patches of the solvent-exposed capping repeats in (a) and (d). The Figure and the charge calculations were made using MOLMOL.[54]

The C-terminal capping repeat (Figure 4) consisted of amino acid residues 129 to the end of 1AWC_B. However, the β-turn had to be adapted to our design, and was changed from the sequence SKFC to DKFG. Ser was replaced by Asp to fit the consensus, and Cys was changed to Gly as in the consensus to prevent problems in oxidizing environments. The chosen sequence was: VNAQ DKFGKT AFDISIDNGN EDLAEILQ.

PHD secondary structure prediction[31] of a construct consisting of the N-terminal capping AR, three AR modules and the C-terminal capping AR was in accordance with our design. For this analysis, the randomized positions and position 26 of the AR modules were not defined (i.e. residues submitted as x). In a prediction of degradation by scanning for PEST[32] sequences and by an analysis with PEPTIDESORT of the GCG package,[33] a construct consisting of the N-terminal capping AR, three AR modules and the C-terminal capping AR showed results equivalent to GABPβ1 (in this case, the randomized positions and position 26 of the AR modules were defined using the corresponding residues of GABPβ1).

Assembly of DNA libraries encoding designed AR domains

The peptide sequences of consensus D (Figure 3) and the N and C-terminal capping repeats were backtranslated into DNA using codons optimal for *Escherichia coli* expression.[33] Multiple copies of a single base were prevented if possible. The codons of the randomized positions (2, 3, 5, 13, 14 and 33) of the designed AR module were encoded by trinucleotides, since they efficiently restrict variability by encoding library positions with a defined mixture of specific base triplets.[34] Using this strategy, we allowed A, D, E, H, K, N, Q, R, S, T with 7% probability each, and F, I, L, M, V, W, Y with 4.3% probability each. The randomized framework position 26 was defined by the degenerate codon HAC, which codes for His, Tyr or Asn.

The modular structure of repeat domains suggests assembling them in a stepwise fashion. Hence, the capping repeats and the designed AR module were assembled separately. The constant N and C-terminal capping repeats were PCR-assembled and subcloned individually into pPANK (see Materials and Methods). The designed AR modules were PCR-assembled and subcloned for sequence analysis; five of eight modules showed no error. To construct DNA cassettes encoding whole AR domains, PCR-assembled designed repeat modules were ligated stepwise to the previously PCR-assembled N-terminal capping repeat by using type IIs restriction enzymes (Figure 5). By this strategy, DNA pools encoding the N-terminal capping repeat and two (N2), three (N3) or four (N4) designed AR modules were obtained. These ligation products were then cloned into a vector containing the cloned C-terminal capping repeat to obtain DNA encoding full-length proteins (Figure 5). The full-length proteins were termed N2C, N3C and N4C, reflecting their content of two, three or four repeat modules, respectively, between the N and the C-terminal capping repeats (resulting in four, five or six repeats in total).

With the seven randomized positions per designed AR module, the theoretical diversity amounts to $3 \times 17^6 = 7.2 \times 10^7$ per repeat. The N2C and N3C libraries will thus have theoretical diversities of $(3 \times 17^6)^2 = 5.2 \times 10^{15}$ and $(3 \times 17^6)^3 = 3.8 \times 10^{23}$, respectively.

Sequence analysis of unselected library members

Having cloned libraries of AR proteins of distinct repeat numbers, we wanted to assess their quality at the DNA level. Analysis of single library members should reveal possible sequence bias. DNA sequencing showed that eight of 14 N2C constructs, six of 19 N3C and four of 19 N4C were correct at the DNA level (i.e. no frameshift, no stop codon, correct framework residue codons and correct trinucleotide codons). The percentage of correct clones decreased with increasing repeat number, as expected. Sequencing of 28 erroneous constructs revealed that 11 errors were located in the N-terminal capping AR, which was generated by assembly PCR using standard oligonucleotides. This high error rate is probably due to the lower quality of these standard oligonucleotides compared to the oligonucleotides containing trinucleotide mixtures. The remaining mutations were located in the designed repeat modules. Four of the 28 errors resulted in frameshifts.

Each designed repeat module of 33 amino acid residues contained six randomized positions, which were encoded by trinucleotides.[34] In total, 777 randomized positions containing trinucleotides were sequenced, showing an approximately random distribution of the codons. As in the leucine-rich repeat protein libraries used by Stumpp *et al.*,[13] Asn was overrepresented (12.1% *versus* 7% expected). Glu, Gln, Arg and Trp were underrepresented (3.1%, 2.3%, 4.4% and 2.3% found *versus* 7%, 7%, 7% and 4.3% expected, respectively). The other codons were found at a frequency less than 2% different from the expected value (see Materials and Methods). Twelve mutations were observed that were not encoded by trinucleotides (one amber, two Gly, nine other amino acids), but were most probably accumulated during the extensive PCR. Apart from these, no undesired codon (Cys, Gly, Pro, stop) was found in the trinucleotide positions. In rare cases (0.4%), entire trinucleotides were missing. The seventh randomized position, consensus framework position 26, was occupied by 30% His, 30% Asn and 40% Tyr (128 positions sequenced). Hence, no clear sequence bias was detectable at any randomized

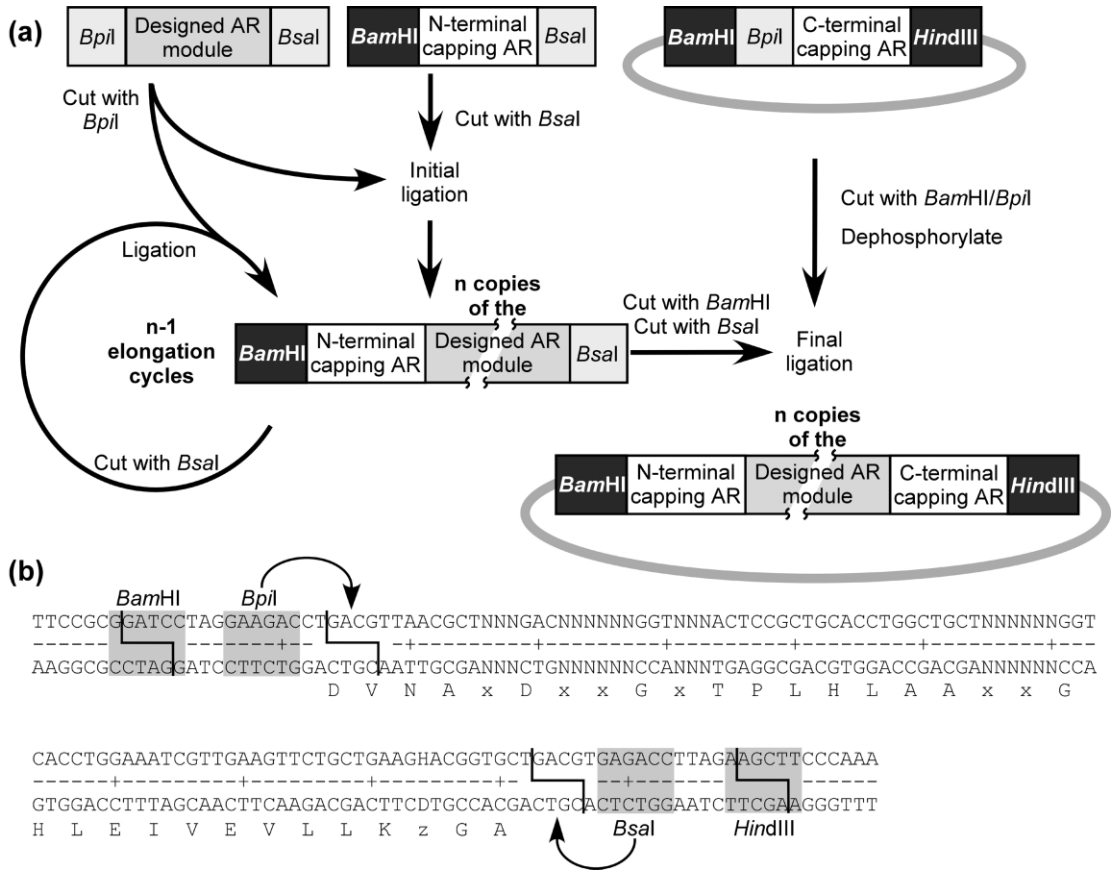


Figure 5. Assembly of designed AR domains at the DNA level and DNA sequence of a designed AR module. (a) The N-terminal capping AR and the designed AR modules are generated by assembly PCR. The N-terminal capping AR is ligated to the first AR module using the type IIs restriction enzymes *Bpi* I and *Bsa* I. To the resulting N1 molecule, more AR modules can be ligated step by step, yielding N2, N3, N4 and longer molecules. Once the desired number of AR modules is connected to the N-terminal capping AR, the construct can be cloned into a vector containing the C-terminal capping AR. By this strategy, AR domains of N1C, N2C, N3C, N4C and longer can be generated. The use of type IIs restriction enzymes ensures the seamless junction of the repeats in a directional manner. Type II restriction sites are represented by black boxes, type IIs restriction sites by light grey boxes. The AR module is represented as a grey box, the N and C-terminal ankyrin capping modules as white boxes. (b) The assembly PCR product of a single AR module is shown. The restriction enzyme recognition sites are shown as grey boxes and the cutting sites are indicated with continuous lines. Note that the DNA recognition sites of the type IIs restriction enzymes are distant from their cleavage site, and thus these sites are lost upon cleavage.

position. Importantly, 75% of the designed AR modules were correct.

Biophysical characterization of randomly chosen library members

We wanted to validate both our strategy and our AR library design by the biophysical analysis of unselected library members, i.e. randomly chosen constructs with correct DNA sequences. The analysis consisted of expression and solubility tests, CD spectroscopy and thermal denaturation. Furthermore, equilibrium unfolding and crystallography were performed.[29]

A first expression screening revealed that all of the above library members that were correct at the DNA level could indeed be expressed in soluble form in large amounts in *E. coli*. The corresponding proteins ran at the expected molecular mass position during SDS-15% PAGE. Six of the correct clones, named E2_5 and E2_17 (N2C library members), E3_5 and E3_19 (N3C library members) and E4_2 and E4_8 (N4C library members) were chosen randomly to be analyzed further. Expression at 37 °C (Figure 6) yielded up to 200 mg/l of soluble protein. Immobilized metal-ion affinity chromatography (IMAC) purification yielded pure protein in a single step, as judged from SDS-15% PAGE (Figure 6). The molecular mass values of the proteins were confirmed by mass spectroscopy. At 10 mg/ml in TBS_{150} (pH 8.0; see Materials and Methods), the proteins remained soluble and did not aggregate over several weeks at 4 °C. An IMAC purification of E3_5 gave sufficiently pure material to successfully determine its structure by X-ray crystallography (Figure 2).[29]

Following SDS-15% PAGE, additional bands were occasionally observed with slightly higher or lower apparent molecular mass than expected. Gel-filtration, mass spectroscopy (data not shown)

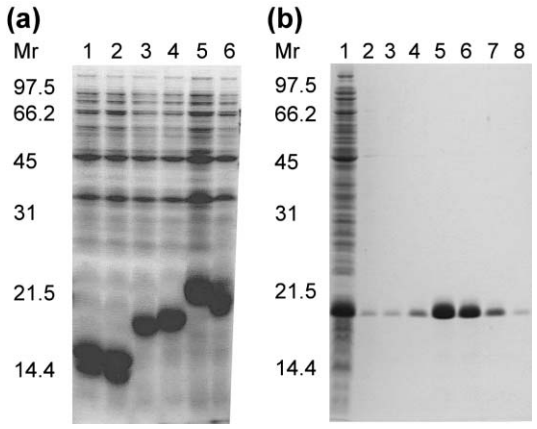


Figure 6. Expression and purification of unselected AR protein library members. (a) Crude extracts of *E. coli* XL1-Blue expressing six consensus AR proteins (see Materials and Methods). Proteins were expressed for four hours and the cell lysates were analyzed by SDS-15% PAGE (lane 1, E2_5; 2, E2_17; 3, E3_5; 4, E3_19; 5, E4_2; 6, E4_8). (b) Single-step IMAC purification of E3_5, an unselected N3C library member (lane 1, column flow-through of the overloaded column; 2, last 1 ml of column wash; 3–8, elution fractions). The size marker is indicated in kDa.

and dynamic light-scattering[29] could not confirm the presence of any protein species other than the expected one. Extensive boiling in SDS and SDS/urea buffer did, however, change the pattern and the relative band intensity in SDS-PAGE analyses,

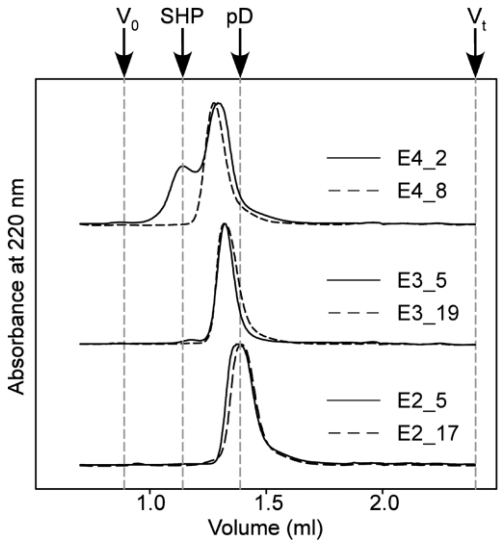


Figure 7. Size-exclusion chromatography of designed AR proteins. The chromatograms of N2C (E2_5 and E2_17), N3C (E3_5 and E3_19) and N4C (E4_2 and E4_8) molecules are shown. All molecules are monomeric, except E4_2, which is a mixture of monomer and (presumably) dimer. The void volume ($V_0 = 0.95$ ml), the total volume ($V_t = 2.4$ ml) and the molecular mass standards (phage protein D with an apparent mass of 17.6 kDa; phage protein SHP, a trimer with an apparent mass of 50.2 kDa)[55] are indicated by broken gray lines in the graph.

Table 1. Table 1 Biophysical data of designed AR proteins of varying length

Protein	CD_{222} (MRE)[a]	MW_{calc} (kDa)[b]	MW_{obs} (kDa)[c]	T_m (°C)[d]	ΔG (kcal/mol)[e]
E2_5	−11,600	14.4	19	79	11.4 ± 0.7
E2_17	−10,300	14.4	18	70	9.5 ± 0.6
E3_5	−11,300	17.7	23	>85	14.8 ± 2.0
E3_19	−12,000	17.8	24	66	9.6 ± 0.5
E4_2[f]	−9400	21.2	30	85	–
E4_8	−11,900	21.3	29	79	21.1 ± 1.3

[a] Mean residue ellipticity (deg cm^2 dmol^{-1}) at 222 nm.
[b] As calculated from the sequence.
[c] As determined by gel-filtration.
[d] As determined by thermal unfolding observing the CD signal at 222 nm.
[e] Data from Ref. 29.
[f] E4_2 is a mixture between a monomer and (presumably) a dimer. The monomer value is listed in MW_{obs}.

suggesting that the multiple bands correspond to different conformers, which are stable during SDS-PAGE.

Size-exclusion chromatography showed that five of the six designed AR proteins were monomeric, and only a single protein species was observed (Figure 7; Table 1). The sixth protein, E4_2, could be purified as a monomer. However, it turned into a mixture of monomer and (presumably) dimer over time at 4 °C (Figure 7). The molecular mass values obtained from the gel-filtration studies are given in Table 1. The observed molecular mass is always slightly higher (by a factor of 1.25 to 1.42) than the calculated value, which can be interpreted to reflect the elongated shape of AR domains in combination with a flexible N-terminal tail (MRGS-HHHHHHGS), which leads to increased hydrodynamic radii of the molecules. In addition, the monomeric state of E3_5 was confirmed by its

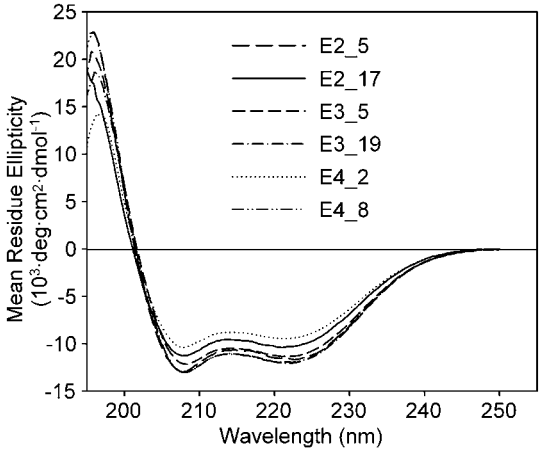


Figure 8. Circular dichroism spectra of designed AR proteins. The spectra of two unselected members from the N2C (E2_5 and E2_17), N3C (E3_5 and E3_19) and N4C (E4_2 and E4_8) library are shown. All proteins exhibit spectra and α-helical content identical with those of natural AR proteins.

crystal structure.[29] We therefore conclude that the majority of the proteins are stable monomers.

The CD spectra of the IMAC-purified protein samples were determined (Figure 8; Table 1). The recorded spectra can be superimposed on the spectra of natural AR proteins such as myotrophin,[35] notch,[36] p19,[37] and an engineered form of p16, p16(Δ1-8)-His.[38] The secondary structure composition of our designed AR proteins thus corresponds to natural AR proteins. In the case of E4_2, oligomerization may influence the spectrum. Combining the CD data with the findings from the protein expression and gel-filtration experiments, we conclude that the designed AR proteins form soluble, monomeric domains having an AR domain fold as designed. For E3_5, the AR domain fold was confirmed by X-ray crystallography (Figure 2).[29]

Thermal stability

To assess the thermal stability of the randomly chosen AR protein library members, heat denaturation was measured by observing the CD signal at 222 nm. All proteins showed cooperative unfolding while exhibiting considerable heat resistance (Figure 9; Table 1). The midpoints of the cooperative transitions in physiological buffer were between 66 °C and more than 85 °C (Table 1). For E4_2, a discontinuity in the CD signal in the pre-transition baseline was observed, probably due to a shift of the monomer/dimer mixture towards a single molecular species. The midpoint of denaturation of E3_5 could not be determined, since the post-transition baseline was not reached after heating the sample to 95 °C. The heat denaturation was only partly reversible for all proteins. The observed high degree of thermal stability reflects the high degree of thermodynamic stability of the designed AR proteins.[29] In GdmCl equilibrium unfolding experiments, the proteins showed cooperative unfolding with midpoints from 2.9 M to 5.1 M GdmCl. Assuming two-state unfolding, ΔG values of unfolding from 9.5 kcal/mol to 21.1 kcal/mol were calculated (Table 1).[29]

Module-wise elongation of AR domains

The N3C library member E3_5 was used to demonstrate the feasibility of a module-wise elongation of AR domains by single repeats. Using PCR cloning, we elongated the N3 part of E3_5 to N5C and N6C domains. Again, the corresponding proteins were expressed in large amounts and could be purified in a single IMAC purification step. Moreover, the proteins exhibited CD spectra identical with those of the other designed AR proteins. In size-exclusion chromatography, two N5C proteins were monomeric with an apparent molecular mass of 39.2 kDa (expected 24.7 kDa) and 37.7 kDa (expected 24.7 kDa), respectively. One N6C protein was monomeric with an apparent molecular mass of 42.7 kDa (expected 28.2 kDa). One N5C protein was a mixture between oligomers and a monomer with an apparent molecular mass of 38.1 kDa (expected 24.7 kDa). One N6C protein had a monomer peak at 44.2 kDa (expected 28.2 kDa) but showed some aggregation.

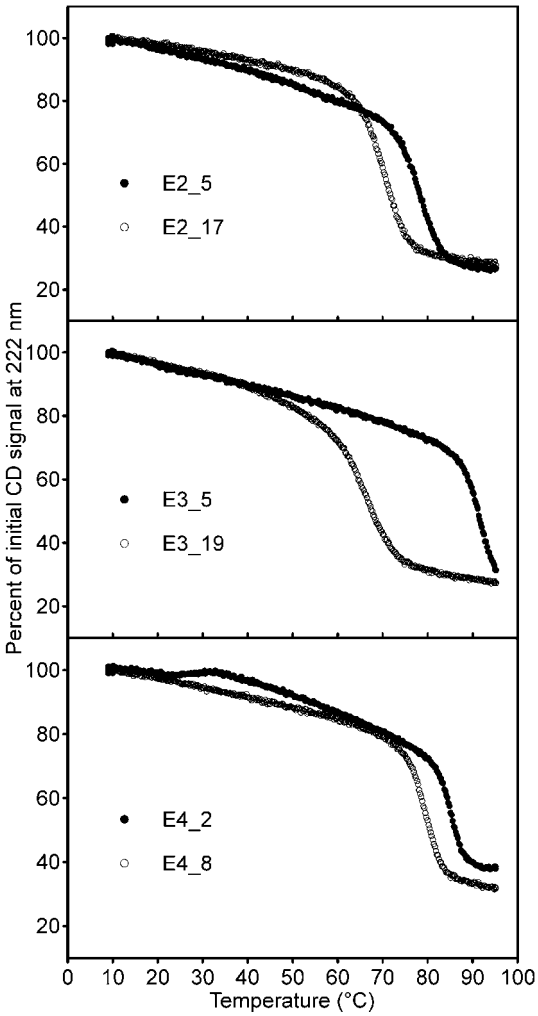


Figure 9. Thermal denaturation of designed AR proteins. Six unselected members of designed AR protein libraries were measured, two each from the N2C (E2_5 and E2_17), N3C (E3_5 and E3_19) and N4C (E4_2 and E4_8) libraries. The denaturation was monitored by observing the CD signal at 222 nm (see Materials and Methods). The CD signal is represented as a percentage of the initial CD signal at 10 °C. Note that this representation makes no assumption about the pre-transition or post-transition baseline.

Discussion

The designed AR proteins possess very favorable biophysical properties

We have developed a novel strategy for constructing combinatorial repeat protein libraries.[3] Here, we have applied this strategy to AR proteins. The accompanying paper by Stumpp *et al.*[13] shows

the application of this strategy to leucine-rich repeat proteins. Using the strategy described here, we were able to generate combinatorial libraries of AR proteins of varying length. The analysis of unselected library members revealed that our design leads to AR proteins with very favorable biophysical properties. We focused our analysis on N2C, N3C and N4C library members, since most natural AR proteins possess repeat numbers in this range. Like natural AR proteins, our designed AR proteins can be expressed in large amounts (Figure 6) but, while the expression of natural AR proteins often results in the formation of inclusion bodies, our designed proteins are expressed in soluble form in the cytoplasm of *E. coli* and remain soluble and folded over weeks at 4 °C. The designed AR proteins are monomeric (Figure 7; Table 1) and, as indicated by CD spectroscopy, they exhibit secondary structure compositions indistinguishable from those of natural AR proteins (Figure 8; Table 1). The crystal structure of the N3C library member E3_5 was determined and it was shown that the designed protein has an AR domain fold.[29]

In thermal denaturation, unselected library members showed cooperative unfolding with mid-points of the cooperative transition ranging from 66 °C to above 85 °C (Figure 9; Table 1). Natural AR proteins that have been tested denature around or below 50 °C, as indicated by CD measurements of notch variants (N4C and N5C) and myotrophin (N2C).[35,39] Thermal denaturation midpoints depend very much on experimental conditions such as protein concentration, buffer composition and the temperature ramp, i.e. the kinetics of unfolding and aggregation. However, the differences in stability between natural and designed AR proteins are large enough to indicate that our designed AR proteins are considerably more stable. The thermal denaturation data reflect the high-level thermodynamic stabilities measured by denaturant-induced equilibrium unfolding of the unselected AR proteins presented here (Table 1).[29]

We were able to design a self-compatible AR module, which could be cloned in various numbers between designed capping ARs, leading to well-expressed, soluble, folded and stable AR domains.

Consensus design of AR proteins

Besides its importance for the self-compatibility of AR modules, our consensus design resulted in very stable AR proteins (Figure 9 and Table 1).[29] These results are consistent with effects of previous consensus design approaches. Consensus strategies have been used to generate enzymes with improved thermostability[40] and to improve antibody stability.[41–43] Stability is not a main selection criterion in the evolution of proteins, once a threshold stability is reached that allows the protein to fulfil a function.[44] The most stable variants may not necessarily be implemented by naturally occurring sequences, but they may be encoded by consensus or "canonical" sequences.[43] We used extensive structural criteria to finally decide on the consensus sequence. We observed a remarkable gain in stability, suggesting that the juxtaposition of AR modules had a synergistic effect.

When examining the crystal structure of E3_5 (Figure 2), we were able to pinpoint several features of the designed AR proteins that could explain this increased stability, such as the absence of irregularities or the presence of extended H-bond networks.[29] These stability findings are similar to data published recently by Mosavi *et al.*,[45] who analyzed full consensus AR proteins. These full consensus AR proteins are based on a slightly different consensus sequence compared to ours. Differences in the two consensus sequences are mainly at positions where they considered sequence data for consensus definition, while we used structural decision parameters (framework residue positions Val19, Lys21, Leu22, Glu25, Ala26 vs. Ile19, Glu21, Val22, Lys25, and a mixture of His, Asn and Tyr at position 26 in our molecules; see Results). The structures of their full consensus proteins are nearly identical with the structure of our designed ARs in E3_5 ($RMSD_{C\alpha}$ 0.51 Å compared to PDB entry 1N0Q). Another important difference is the presence of capping repeats (Figure 4) sealing the stack of designed ARs in our molecules compared to the full consensus AR proteins described by Mosavi *et al.*[45] Experimentally, the proteins described by Mosavi *et al.*[45] are more soluble at acidic pH (pH 4–5) than at neutral pH, while our proteins are soluble and stable under physiological conditions (20–50 mM Tris–HCl (pH 6.5–8.5), 50–500 mM NaCl). Both the differences in consensus sequence and the presence of capping ARs in our constructs might lead to this altered behavior. Nevertheless, both studies show that the AR framework is intrinsically very stable. This stability could, in part, account for the abundance of AR proteins in nature. Main and co-workers[46] recently reported the consensus design of tetratricopeptide repeat proteins of different repeat numbers. They also observed high thermal stability of the consensus designed proteins.

Module-wise assembly of AR proteins

The modular nature of AR proteins suggests assembling repeat domains module-wise. We used type IIs restriction enzymes for this purpose, which allow the cloning of repeats in a directional manner, independent of the repeat sequence (Figure 5). The advantage of type IIs restriction enzymes is the freedom of choice of the ligation site. It could, in principle, be placed in any part of the repeat module except for the randomized positions. We did not want to affect the capping repeat structure and sequence; therefore, the ligation site could not be placed in any of the α-helices or in

the short loop connecting the helices. The only remaining possibility was the loop connecting the second α-helix with the β-turn of the next repeat.

Similar to the approach described in the accompanying paper,[13] the chosen cloning strategy using type IIs restriction enzymes (Figure 5) allows a number of evolution strategies that are amenable only to repeat proteins. For example, repeats might be shuffled, added or subtracted. In nature, the IκBα/Bcl-3 pair gives an example of such a repeat extension, where the repeat number reflects the different binding properties of these molecules.[47] We have shown the feasibility of this extension approach for the designed AR proteins by elongating E3_5 (N3C) to N5C or N6C.

In addition to repeat shuffling and module-wise addition or subtraction of single repeats, other evolution strategies are amenable to our molecules. Alterations from the 33 amino acid residue consensus, similar to what is observed in single repeats of IκBα,[23,24] Swi6[28] or the INK4 family members, could be used for the improvement of selected binding molecules *via in vitro* evolution or rational design. Another evolution strategy could involve increasing or decreasing the AR domain curvature, which can be achieved by varying specific framework residues.[10]

Designed AR proteins in biotechnology

Our findings show that we have libraries of well-behaved AR proteins for use in selection procedures. The proteins are expressed in large amounts, they are soluble, monomeric and stable under physiological conditions, they are cysteine-free and allow a great variety of amino acids in the randomized positions. Therefore, these molecules exactly match the requirements for novel scaffolds to be used for the generation of novel binding proteins. The stability of the consensus designed AR proteins is sufficiently high that some losses in stability can be tolerated during the course of directed evolution of the designed AR proteins.

In our AR domains, the potential target interaction residues are located in the β-turn and the first α-helix of each repeat module, creating a large and modular interaction surface (Figure 2).[3,29] This interaction mode extends and combines previous concepts in the field of combinatorial libraries. Usually, either flexible loops (e.g. Knappik *et al.*[41]) or rigid, flat surfaces (e.g. Nord *et al.*[48]) were randomized, but not a combination of turns and helices, which constitute a continuous surface that can be extended by adding more repeats.

We have recently used designed N2C and N3C AR protein libraries in ribosome display[49,50] selections against various globular proteins. Specific nanomolar binders were obtained, which prove the success of designed AR proteins as novel scaffolds for molecular recognition. A detailed analysis of these experiments will be published elsewhere. Because of their favorable biophysical properties, designed AR proteins could ideally serve as recognition molecules on protein chips. Similarly, the absence of intracellular aggregation or misfolding and the absence of cysteine residues would allow these proteins to be used as intracellular protein binders or enzyme inhibitors. In this regard, designed AR proteins could be an attractive and more stable alternative to intrabodies.[51]

Conclusions

We have successfully implemented our novel strategy harnessing the modular nature of repeat proteins for the generation of designed AR protein libraries. Through sequence and structure consensus analyses, we designed an AR module composed of fixed framework positions and randomized potential interaction positions. AR domains were generated by cloning two, three or four designed modules between N and C-terminal capping repeats. All tested proteins exhibit very favorable biophysical properties. They can be expressed in soluble form in large amounts and they can be purified easily. They are monomeric and show CD spectra indistinguishable from those of natural AR proteins. Furthermore, they are exceptionally resistant to heat denaturation. These findings suggest that the abundance of natural AR proteins is, at least in part, based on the exceptional properties of the AR framework, a stable and modular protein–protein interaction motif. Our findings show that we can build modular and stable proteins with randomized surfaces that may be used to create novel binding molecules. The modular structure of repeat proteins will allow completely new evolution strategies that are not feasible with classical scaffolds.

Material and Methods

In silico analysis

We used the SMART†,[7] the GenBank‡,[18] and the PDB§ [52] databases for our analyses. Clustal W‖ [14] and BLAST¶ [17] were used for alignments. Structural modeling was done with InsightII (Accelrys, USA). NACCESS helped to identify target interaction residues from structures of complexes. PHD prediction[a] [31] was used for secondary structure prediction. PEST[b] [32] and PEPTIDESORT of GCG (Accelrys, USA)[33] were used to

† http://smart.embl-heidelberg.de
‡ http://www.ncbi.nlm.nih.gov/Genbank/
§ http://www.pdb.org
‖ http://www.ch.embnet.org/software/ClustalW.html
¶ http://www.ncbi.nlm.nih.gov/blast/
a http://cubic.bioc.columbia.edu/predictprotein/
b http://www.at.embnet.org/embnet/tools/bio/PESTfind/

compare designed and natural AR proteins. GCG was used for designing the DNA sequence.

General molecular biology

Unless stated otherwise, all experiments were performed as described.[53] Enzymes and buffers were from New England Biolabs (USA) or Fermentas (Lithuania). The cloning and production strain was *E. coli* XL1-Blue (Stratagene, USA). The cloning and protein expression vector was pPANK, a pQE30 (QIAgen, Germany) derivative lacking the *Bbs* I and *Bsa* I sites. pPANK was generated *via* PCR-cloning using the oligonucleotides Bbs I (5′-TGATTTCTCGAGGTGTAGTCGAAAGGGCCTCGTG-3′), Bsa I (5′-GCAATGATACCGCGAGAACCACGCTCA CCGGC-3′) and Avr2 (5′-CCGCCGCTCTAGAGGGAAA CCTAGGGCTGCCTCGCGCG-3′) and pQE30 as template. The oligonucleotides *Bbs* I and *Bsa* I were used to generate a PCR product, which was used as primer in a second PCR reaction together with oligonucleotide Avr2. The resulting PCR product was *Xho* I/*Xba* I-digested and ligated to the *Xho* I/*Xba* I promoter fragment of pQE30.

Synthesis of DNA encoding AR proteins

Oligonucleotides incorporating mixed trinucleotides as building blocks[34] were from MorphoSys AG (Germany). INT1:

5′-CTGACGTTAACGCTNNNGACNNNNNNGGTN NNACTCCGCTGCACCTGGC-3′ and INT2:

5′-ACTCCGCTGCACCTGGCTGCTNNNNNNGGTC ACCTGGAAATCG-3′.

NNN represents a mixture of trinucleotides encoding the amino acids A, D, E, H, K, N, Q, R, S, T (7% each) and F, I, L, M, V, W, Y (4.3% each). Standard oligonucleotides were from Microsynth (Switzerland).

INT3: 5′-AACGTCAGCACCGTDCTTCAGCAGAAC TTCA ACGATTTCCAGGTGACC-3′; D represents any of the nucleotides A, G or T).

INT4: 5′-AGCAGCCAGGTGCAGCGGAGT-3′.

INT5: 5′-TTCCGCGGATCCTAGGAAGACCTGACGT TAAC GCT-3.

INT6: 5′-TTTGGGAAGCTTCTAGAAGACAACGT CAGCAC CGT-3′.

INT6a: 5′-TTTGGGAAGCTTCTAAGGTCTCACGT CAGCAC CGT-3′.

INT6b: 5′-TTTGGGAAGCTTCTAAGGTCTC-3′.

EWT1: 5′-TTCCGCGGATCCGACCTGGGTAA GAAACTGCT GGAAGCTGCTCGTGCTGGTCAGGAC GACGAAG-3′.

EWT2: 5′-AACGTCAGCACCGTTAGCCATCAGGA TACGAA CTTCGTCGTCCTGACC-3′.

EWT3: 5′-TTCCGCGGATCCGACCTGGG-3′.

TEN3: 5′-TTCCGCGGATCCG-3′.

WTC1: 5′-CTGACGTTAACGCTCAGGACAAATTCG GTAAG ACCGCTTTCGACATCTCCATCGACAACGG TAACGA GG-3′.

WTC2: 5′-TTGCAGGATTTCAGCCAGGTCCTCGT TACCGTT GTC-3′.

WTC3: 5′-TTTGGGAAGCTTCTATTGCAGGATTTCA GC-3′.

The AR modules were generated by assembly PCR using oligonucleotides INT1, INT2, INT3, INT4, INT5 and INT6a, and Vent® Polymerase (one minute annealing at 50 °C; standard buffer with a final concentration of 5.5 mM $MgSO_4$). A subset of the resulting PCR product was cloned *via* *Bam* HI/*Hin*dIII into pPANK and sequenced using standard techniques. The AR module sequence is shown in Figure 5.

The N-terminal capping AR was prepared by assembly PCR using oligonucleotides EWT1, EWT2, TEN3 and INT6. The resulting DNA was cloned *via* *Bam* HI/*Hin*dIII into pPANK. The DNA sequence was verified using standard techniques. The C-terminal capping AR was prepared similarly, but by using oligonucleotides WTC1, WTC2, WTC3 and INT5.

The ligation of the DNA encoding an AR protein from single AR modules and AR capping repeats is represented schematically in Figure 5. To clone DNA encoding AR proteins, the PCR-assembled N-terminal capping AR (using oligonucleotides EWT1, EWT2, TEN3 and INT6a) was cut with *Bsa* I and ligated to a *Bpi* I-cut AR module. The ligation product, termed N1 (where N denotes the N-terminal capping repeat and the digit is the number of randomized repeat modules), was PCR-amplified using oligonucleotides EWT3 and INT6b. The amplified product was cleaved again with *Bsa* I. The subsequent ligation to *Bpi* I-cut AR modules started a new cycle of elongation, which was repeated until the desired number of AR modules was added to the N-terminal capping AR (termed N2, N3, N4 etc.). DNA corresponding to PCR-amplified N2, N3 and N4 were then cut with *Bam* HI/*Bsa* I and ligated to a *Bam* HI/*Bpi* I-cut pPANK containing the C-terminal capping AR (Figure 5). This yielded cloned DNA molecules encoding N2C, N3C and N4C AR protein libraries (where N denotes the N-terminal capping repeat, the digit is the number of randomized repeat modules and C is the C-terminal capping repeat).

An unselected N3C library member (named E3_5, see below) was used as template for repeat protein elongation. Using the oligonucleotides EWT3 and INT6a, fragments corresponding to N, N1, N2 and N3 were generated by PCR. The N3 fragment was isolated and then reamplified using oligonucleotides EWT3 and INT6b. Elongation of the N3 to N5 and N6 fragments and cloning to N5C and N6C molecules was carried out as described above.

Screening for protein expression and DNA sequencing

An SDS-15% PAGE screening of single unselected clones was performed using 10 ml cultures. A stationary overnight culture (5 ml of LB, 1% (w/v) glucose, 100 mg/l of ampicillin; 37 °C) was used to inoculate the cultures (1 ml of inoculum in 9 ml of the above medium). After one hour, protein expression was induced using 200 μM IPTG and cultures were incubated for five hours. In parallel, all screened clones were subjected to DNA sequence analysis. Two unselected clones of the libraries N2C (clone names: E2_5 and E2_17), N3C (E3_5 and E3_19) and N4C (E4_2 and E4_8) were chosen for subsequent protein analyses.

Protein expression and purification

The N2C, N3C, N4C, N5C and N6C clones were expressed as follows: 25 ml of stationary overnight cultures (LB, 1% glucose, 100 mg/l of ampicillin; 37 °C) were used to inoculate 1 l cultures (same medium). At $A_{600} = 0.7$, the cultures were induced with 300 μM IPTG and incubated for four hours. Samples were analyzed by SDS-15% PAGE (Figure 6). The cultures were

centrifuged and the resulting pellets were resuspended in 40 ml of TBS_{500} (50 mM Tris–HCl (pH 8.0), 500 mM NaCl) and sonicated. The lysate was recentrifuged and glycerol (10% final concentration) and imidazole (20 mM final concentration) were added to the resulting supernatant. Proteins were purified over a Ni-nitrilotriacetic acid column (2.5 ml column volume) according to the manufacturer's instructions (QIAgen, Germany; Figure 6).

Size-exclusion chromatography

IMAC-purified proteins were analyzed on a Superdex 75 gel-filtration column (Amersham Pharmacia Biotech, USA) using a Pharmacia SMART system at a flow-rate of 60 μl/minute and with TBS_{150} (50 mM Tris–HCl (pH 7.4), 150 mM NaCl) as running buffer (Figure 7).

CD spectroscopy

Circular dichroism spectra were recorded in 10 mM sodium phosphate buffer (pH 6.5), 100 mM NaCl, using 10 μM purified protein on a Jasco J-715 instrument (Jasco, Japan). The CD signal was converted to mean residue ellipticity using the concentration of the sample determined spectrophotometrically at 280 nm under denaturing conditions (Figure 8).

Heat denaturation was performed in 20 mM sodium phosphate (pH 7.4), 200 mM NaCl with 10 μM protein and a temperature shift from 10 °C to 95 °C within 120 minutes. CD data were collected at 222 nm every 20 seconds with a bandwidth of 2 nm and 16 seconds response time (Figure 9).

Data Bank accession numbers

The DNA and amino acid sequences of proteins E2_5, E2_17, E3_5, E3_19, E4_2 and E4_8 have been deposited in GenBank[18] with accession numbers AY195851, AY195852, AY195853, AY195854, AY195855 and AY195856, respectively. The DNA sequence of pPANK has been deposited in GenBank[18] (accession number AY327140).

Acknowledgements

We thank the members of the Plückthun laboratory for valuable discussions, Dr Annemarie Honegger for EXCEL® macros as well as Dr David L. Zechel and Dr Casim A. Sarkar for critical reading of the manuscript. We thank MorphoSys AG for the trinucleotide-containing oligonucleotides. H.K.B. was supported by a pre-doctoral fellowship of the Roche Research Foundation. M.T.S. was the recipient of an FCI and a BMBF pre-doctoral scholarship. This work was supported by the Swiss National Centre of Competence in Research (NCCR) in Structural Biology and the Swiss Cancer Research grant KFS 1055-09-2000.

References

1. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001). Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131.
2. Kobe, B. & Kajava, A. V. (2000). When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515.
3. Forrer, P., Stumpp, M. T., Binz, H. K. & Plückthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Letters*, **539**, 2–6.
4. Nygren, P.-Å. & Uhlén, M. (1997). Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* **7**, 463–469.
5. Skerra, A. (2000). Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167–187.
6. Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins: Struct. Funct. Genet.* **17**, 363–374.
7. Letunic, I., Goodstadt, L., Dickens, N. J., Doerks, T., Schultz, J., Mott, R. *et al.* (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucl. Acids Res.* **30**, 242–244.
8. Gorina, S. & Pavletich, N. P. (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science*, **274**, 1001–1005.
9. Sedgwick, S. G. & Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316.
10. Michaely, P., Tomchick, D. R., Machius, M. & Anderson, R. G. (2002). Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J.* **21**, 6387–6396.
11. Suzuki, F., Goto, M., Sawa, C., Ito, S., Watanabe, H., Sawada, J. & Handa, H. (1998). Functional interactions of transcription factor human GA-binding protein subunits. *J. Biol. Chem.* **273**, 29302–29308.
12. Malek, S., Huxford, T. & Ghosh, G. (1998). IκBα functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF-κB. *J. Biol. Chem.* **273**, 25427–25435.
13. Stumpp, M. T., Forrer, P., Binz, H. K. & Plückthun, A. (2003). Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.* **332**, 471–487.
14. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
15. Breeden, L. & Nasmyth, K. (1987). Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of *Drosophila*. *Nature*, **329**, 651–654.
16. Lux, S. E., John, K. M. & Bennett, V. (1990). Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature*, **344**, 36–42.
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
18. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2000). GenBank. *Nucl. Acids Res.* **28**, 15–18.
19. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D. & Sonnhammer, E. L. (1999). Pfam 3.1: 1313

multiple alignments and profile HMMs match the majority of proteins. *Nucl. Acids Res.* **27**, 260–262.
20. Luh, F. Y., Archer, S. J., Domaille, P. J., Smith, B. O., Owen, D., Brotherton, D. H. *et al.* (1997). Structure of the cyclin-dependent kinase inhibitor p19Ink4d. *Nature*, **389**, 999–1003.
21. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L. & Wolberger, C. (1998). The structure of GABPα/ankyrin repeat heterodimer bound to DNA. *Science*, **279**, 1037–1041.
22. Byeon, I.-J.L., Li, J., Ericson, K., Selby, T. L., Tevelev, A., Kim, H. J. *et al.* (1998). Tumor suppressor $p16^{INK4a}$: determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. *Mol. Cell*, **1**, 421–431.
23. Huxford, T., Huang, D. B., Malek, S. & Ghosh, G. (1998). The crystal structure of the IκBα/NF-κB complex reveals mechanisms of NF-κB inactivation. *Cell*, **95**, 759–770.
24. Jacobs, M. D. & Harrison, S. C. (1998). Structure of an IκBα/NF-κB complex. *Cell*, **95**, 749–758.
25. Yang, Y., Nanduri, S., Sen, S. & Qin, J. (1998). The structural basis of ankyrin-like repeat function as revealed by the solution structure of myotrophin. *Structure*, **6**, 619–626.
26. Venkataramani, R., Swaminathan, K. & Marmorstein, R. (1998). Crystal structure of the CDK4/6 inhibitory protein $p18^{INK4c}$ provides insights into ankyrin-like repeat structure/function and tumor-derived $p16^{INK4}$ mutations. *Nature Struct. Biol.* **5**, 74–81.
27. Mandiyan, V., Andreev, J., Schlessinger, J. & Hubbard, S. R. (1999). Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-associated protein β. *EMBO J.* **18**, 6890–6898.
28. Foord, R., Taylor, I. A., Sedgwick, S. G. & Smerdon, S. J. (1999). X-ray structural analysis of the yeast cell cycle regulator Swi6 reveals variations of the ankyrin fold and has implications for Swi6 function. *Nature Struct. Biol.* **6**, 157–165.
29. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl Acad. Sci. USA*, **100**, 1700–1705.
30. O'Neil, K. T. & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, **250**, 646–651.
31. Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* **266**, 525–539.
32. Rogers, S., Wells, R. & Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, **234**, 364–368.
33. Womble, D. D. (2000). GCG: The Wisconsin Package of sequence analysis programs. *Methods Mol. Biol.* **132**, 3–22.
34. Virnekäs, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G. & Moroney, S. E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucl. Acids Res.* **22**, 5600–5607.
35. Mosavi, L. K., Williams, S. & Peng, Z.-y. (2002). Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein. *J. Mol. Biol.* **320**, 165–170.
36. Zweifel, M. E. & Barrick, D. (2001). Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor. 1. Solution conformational and hydrodynamic properties. *Biochemistry*, **40**, 14344–14356.
37. Zeeb, M., Rosner, H., Zeslawski, W., Canet, D., Holak, T. A. & Balbach, J. (2002). Protein folding and stability of human CDK inhibitor $p19^{INK4d}$. *J. Mol. Biol.* **315**, 447–457.
38. Zhang, B. & Peng, Z.-y. (2000). A minimum folding unit in the ankyrin repeat protein $p16^{INK4}$. *J. Mol. Biol.* **299**, 1121–1132.
39. Zweifel, M. E. & Barrick, D. (2001). Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor. 2. Solution stability and cooperativity of unfolding. *Biochemistry*, **40**, 14357–14367.
40. Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta*, **1543**, 408–415.
41. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G. *et al.* (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86.
42. Ohage, E. & Steipe, B. (1999). Intrabody construction and expression. I. The critical role of V_L domain stability. *J. Mol. Biol.* **291**, 1119–1128.
43. Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**, 188–192.
44. Wang, Q., Buckle, A. M. & Fersht, A. R. (2000). Stabilization of GroEL minichaperones by core and surface mutations. *J. Mol. Biol.* **298**, 917–926.
45. Mosavi, L. K., Minor, D. L., Jr & Peng, Z.-y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. USA*, **99**, 16029–16034.
46. Main, E. R., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure (Camb.)*, **11**, 497–508.
47. Michel, F., Soler-Lopez, M., Petosa, C., Cramer, P., Siebenlist, U. & Müller, C. W. (2001). Crystal structure of the ankyrin repeat domain of Bcl-3: a unique member of the IκB protein family. *EMBO J.* **20**, 6180–6190.
48. Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M. & Nygren, P-Å. (1997). Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nature Biotechnol.* **15**, 772–777.
49. Hanes, J. & Plückthun, A. (1997). *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA*, **94**, 4937–4942.
50. Amstutz, P., Forrer, P., Zahnd, C. & Plückthun, A. (2001). *In vitro* display technologies: novel developments and applications. *Curr. Opin. Biotechnol.* **12**, 400–405.
51. Cattaneo, A. & Biocca, S. (1999). The selection of intracellular antibodies. *Trends Biotechnol.* **17**, 115–121.
52. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K. *et al.* (2002). The Protein Data Bank. *Acta Crystallog. sect. D*, **58**, 899–907.
53. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

54. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
55. Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E. *et al.* (2000). Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nature Struct. Biol.* **7**, 230–237.

Edited by J. Thornton

(Received 23 April 2003; received in revised form 9 July 2003; accepted 10 July 2003)

Chapter 3

In Vitro Selection with Ribosome Display: Overview and Methodology

Contents

400

In vitro display technologies: novel developments and applications

Patrick Amstutz, Patrik Forrer, Christian Zahnd and Andreas Plückthun*

In vitro display techniques are powerful tools to select polypeptide binders against various target molecules. Novel applications include maturation of protein affinity and stability, selection for enzymatic activity, and the display of cDNA and random polypeptide libraries. Taken together, these display techniques have great potential for biotechnological, medical and proteomic applications.

Addresses
Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
*e-mail: plueckthun@biocfebs.unizh.ch

Current Opinion in Biotechnology 2001, 12:400–405

0958-1669/01/$ – see front matter

Abbreviations
PCR polymerase chain reaction
scFv single-chain Fv

Introduction

The growing interest of the research community and pharmaceutical companies in protein–protein interactions has led to an increasing demand for sophisticated methods for the rapid identification, characterization, and potential improvement of interaction partners. The most popular of these methods, namely the yeast two-hybrid system [1] and phage display [2] (see the article by Sidhu and Pelletier in this issue pp 340–347), are limited by the involvement of living cells in the process of library generation or screening. This is not the case for *in vitro* selection technologies. In these techniques the number of molecules that can be handled is not limited by cellular transformation efficiencies; thus, very large libraries of up to 10^{14} members can be built. This feature also facilitates directed evolution experiments, in which rounds of randomization and selection alternate, as transformation can be avoided between rounds. Furthermore, special reagents such as the reducing agent dithiothreitol or detergents can be added to select binders under conditions chosen by the experimentor.

Two main groups of *in vitro* selection technologies can be distinguished. The first group imitates the compartmentalization of living cells by performing translation and selection within a water-in-oil emulsion [3–5]; this method was recently summarized in an excellent review [6]. This compartmentalization ensures the coupling of genotype and phenotype — a prerequisite for any selection method. The second group, the *in vitro* display technologies, makes use of a physical link between messenger RNA (mRNA) and nascent polypeptide during translation to couple genotype and phenotype. The most popular *in vitro* display technologies are ribosome display and mRNA display (reviewed in [7–12]). This review focuses on the recent advances in the field of *in vitro* display methods and discusses the potential of these technologies for future applications in basic and applied research.

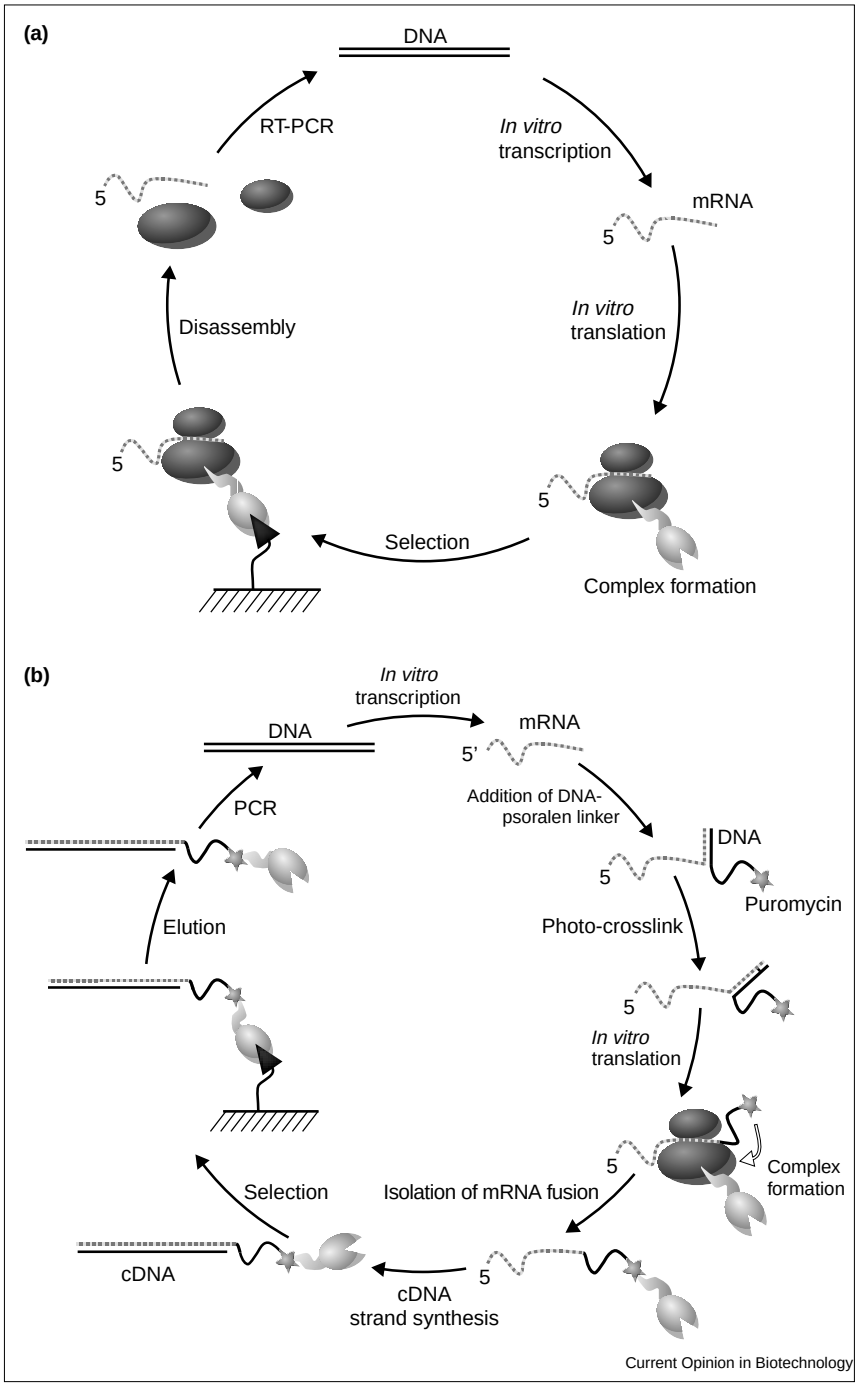
In vitro display technologies

Ribosome display (Figure 1a) was first developed by Mattheakis *et al.* [13] for the selection of peptides and further improved for the selection of folded proteins by Hanes and Plückthun [14] and He and Taussig [15]. This method relies on non-covalent ternary complexes of mRNA, ribosome and nascent polypeptide, ensuring the coupling of genotype and phenotype. A fusion protein is constructed in which the the domain of interest is fused to a C-terminal tether, such that this domain can fold while the tether is still in the ribosomal tunnel. This fusion construct lacks a stop codon at the mRNA level, thus preventing release of the mRNA and the polypeptide from the ribosome. High concentrations of magnesium and low temperature further stabilize the ternary complex. These complexes, which are formed during *in vitro* translation, can directly be used to select for the properties of the displayed protein.

The related technology of mRNA display (Figure 1b), which has also been termed 'mRNA–protein fusions' [16] or '*in vitro* virus' [17], was predominantly developed by Roberts and Szostak. This method relies on the covalent coupling of mRNA to the nascent polypeptide. The mRNA is first covalently linked to a short DNA linker carrying a puromycin moiety. The library is then translated *in vitro*, as in ribosome display. When the ribosome reaches the RNA–DNA junction the ribosome stalls and the puromycin moiety enters the peptidyltransferase site of the ribosome and forms a covalent linkage to the nascent polypeptide. The protein and the mRNA are thus coupled and are subsequently isolated from the ribosome and purified. In the current protocol, a cDNA strand is then synthesized to form a less sticky RNA–DNA hybrid and these complexes are finally used for selection.

The protocol of mRNA display has been significantly improved since it was first reported by extending the method from the display of short peptides to proteins [9]. The authors were also able to increase the yield of functional mRNA–protein fusions about 40-fold compared with the original protocol. Furthermore, the laborious linkage of mRNA to the puromycin-containing DNA linker, thus far an enzymatic ligation reaction, was also significantly improved [18,19]. In the new method, a DNA linker carrying a psoralen moiety is hybridized to the end of the mRNA and directly photo-crosslinked to the mRNA. These improvements have opened the door for mRNA display to handle complex protein libraries, as has been possible with ribosome display [11,15,20,21•,22••,23,24].

Figure 1



In vitro display technologies. (a) Schematic representation of a ribosome display selection round. A DNA library encoding the proteins of interest is genetically fused to a tether, which allows the protein to fold while the tether is still in the ribosomal tunnel. The resulting construct, which lacks a stop codon, is transcribed *in vitro* into mRNA and further translated *in vitro*. The translation is stopped such that stable ternary complexes of mRNA, ribosomes and nascent polypeptides are formed. These complexes are directly used for binding selection on the immobilized target. The mRNA of the bound complexes is rescued by dissociating the ribosome with EDTA. A reverse transcription reaction followed by PCR yields the genetic information of the selected clones. These clones can then be analyzed or used as input for the next selection round. (b) Schematic representation of a mRNA display selection round. A DNA library encoding the proteins of interest is transcribed *in vitro*. The resulting mRNA is covalently fused to a short DNA linker which carries a puromycin moiety at its 3 -end. This linking can be achieved by hybridization and subsequent photo-crosslinking of a psoralen-labeled DNA linker to the mRNA (as shown here) or by an enzymatic ligation reaction (not shown). The resulting construct is translated *in vitro*. During translation the ribosome reaches the RNA–DNA junction and stalls. This allows the puromycin moiety to bind to the ribosomal A site. Thereby, the nascent polypeptide is transferred to the puromycin moiety leading to a covalent mRNA–polypeptide fusion. The mRNA–fusion complex is subsequently purified and the first cDNA strand is synthesized. A selection for binding on immobilized ligand is then carried out and the bound complexes are eluted. The following PCR yields the genetic information of the selected clones. These clones can then be analyzed or used as input for the next selection round. In both (a) and (b), black lines correspond to DNA and dotted lines correspond to mRNA. In (b) the star represents the puromycin moiety. For simplicity, the crosslinked mRNA–DNA hybrid is not shown after the *in vitro* translation step.

The stability of mRNA was repeatedly discussed as a weak point of both ribosome and mRNA display, as RNA is susceptible to hydrolysis and nuclease degradation. Nevertheless, for ribosome display it has been shown that the ternary complexes are stable for up to at least 15 days (C Zahnd *et al.*, unpublished data). To improve the stability of mRNA display, Kurz *et al.* [25•] reported a method to replace the mRNA molecule within the mRNA–protein complex with its double-stranded cDNA. This methodology may be especially attractive for selections under harsh conditions (e.g. high temperatures).

Improved library quality by preselection

The success of selection experiments depends to a large extent on the quality of the library. Although the theoretical size of a library is virtually unlimited, the transformation

efficiency for yeast (10^7–10^8 cells/ g DNA) and for *Escherichia coli* (10^9–10^{10} cells/ g DNA) limits the achievable library size. By contrast, *in vitro* display technologies can handle libraries with up to 10^{14} members, depending only on the scale of the *in vitro* translation used. A common way to generate libraries involves the use of degenerate oligonucleotides; however, such oligonucleotides often contain deletions that result in frame-shifts and the appearance of stop codons, thus decreasing the effective library size. One way to avoid this is to use trinucleotides [26] as building blocks, because even if deletions do occur the correct reading frame is still maintained.

A generally useful approach to enlarge the percentage of correct (i.e. in-frame and full-length) clones in a library is to eliminate problematic sequences through preselection. Because of the large library size accessible in the *in vitro* display technologies, a useful library diversity is maintained even after preselection. The library is cloned between an N- and C-terminal tag and displayed such that selection for the occurrence of these tags will yield in-frame and full-length polypeptides. Premature stop codons cause the ribosome to dissociate, and frame-shifts will alter the C-terminal tag. Nonetheless, suppressor tRNA present in the extract or mistakes in the synthesis machinery appear to still let a fraction of undesired molecules slip through this 'filter'. Cho *et al.* [27•] performed such a preselection using mRNA display. In this way they were able to remove a large fraction of the non-functional proteins from a large-scale *in vitro* translation (10 ml), thus improving the proportion of correct proteins in three different libraries by up to two orders of magnitude. They still maintained a final complexity of about 10^{13} full-length molecules.

Selection experiments with two of these libraries have been published. From one such library, Keefe and Szostak [28••] were able to select polypeptides, presumably with a folded structure, that were able to bind ATP. This library consisted of a completely random stretch of 80 amino acids, which had been preselected as described above. All selected sequences were full-length, indicating that preselection was successful. From their results, the authors estimated the number of ATP binders in a random sequence to be 1 in 10^{11}, which underlines the necessity of having a high-quality full-length library, and a selection method capable of handling libraries of this size.

Using the second of these preselected libraries [27•], consisting of amphipathic -helical or -strand segments, Wilson *et al.* [29] selected for streptavidin binders. The highest affinity of the selected peptides for streptavidin was about 5 nM, in contrast to micromolar affinities of peptides selected in previous phage-display experiments [30]. However, all of the selected peptides were derived from frame-shifted sequences, even though preselection had been performed. Because the library was designed with restricted codon frequencies, such frame-shifted sequences had a 700-fold increased probability of containing the known streptavidin-binding consensus sequence His-Pro-Gln. These results also demonstrate how important the design of the initial library is, as the selected molecules had no similarity to the designed -helical or -stranded elements. Even though preselection was performed, the rare frame-shifted sequences prevailed over the much more abundant in-frame sequences.

Nevertheless, the preselection approach using *in vitro* display technologies may be a powerful tool for improving library quality in terms of enlarging the proportion of correct library members.

Directed evolution of proteins

Natural evolution has efficiently adapted proteins to their tasks under given environmental conditions. Nevertheless, the technological or medical application of proteins often places different demands on them; thus, their performance needs to be optimized. Using the Darwinian principle, evolution of polypeptides can now be conducted in the test tube: a pool of molecules (library) is subjected to alternating rounds of selection and randomization. If the randomization is carried out on the whole library, not only the original library is screened for the best molecules but the library composition is also adapted from round to round so that sequences not present in the original library become accessible to selection. With this approach, Hanes *et al.* [21•] selected, by ribosome display, a set of different antibody single-chain Fv (scFv) fragments from a synthetic naïve library, with up to 40-fold improved affinities when compared with the progenitor sequences present in the original library. All of the selected antibodies had accumulated mutations as a result of amplification with a low-fidelity DNA polymerase. This work demonstrated that protein evolution can be an intrinsic part of each ribosome display cycle. By including an additional diversification step in each round of ribosome display to increase the error rate even further, for example, by error-prone polymerase chain reaction (PCR) [31,32] (described in the article by Kurtzman *et al.* in this issue pp 361–370) or DNA shuffling [33], Jermutus *et al.* [22••] further confirmed the potential of ribosome display for directed *in vitro* protein evolution. They demonstrated that distinct and predictable biophysical characteristics of scFvs, affinity and stability, can be rapidly and efficiently evolved by combining these techniques (see below). Similarly, mRNA display in combination with error-prone PCR proved to be a powerful approach for *in vitro* evolution of proteins as shown by Keefe and Szostak [28••] (see above). Thus, the work of these groups demonstrated that the PCR amplification step inherent to ribosome and mRNA display can be directly exploited for *in vitro* protein evolution by relaxing the accuracy of pool replication during selection cycles. The large library size accessible by these *in vitro* display techniques further supports successful evolution experiments. Taken together, ribosome and mRNA display have great potential for directed protein evolution.

Maturation of protein affinity

Protein affinity maturation with molecular evolution technologies is an important step in producing selective and high-affinity binding proteins for applications in biotechnology and medicine [23]. Both ribosome display and mRNA display have allowed selection for binding proteins to a wide variety of targets, such as small compounds [22••,28••], peptides [20], whole proteins [21•,23,24,34••] or even a specific DNA structure [35••].

Recent work now also demonstrates that *in vitro* display technologies have great potential for the maturation of high-affinity protein binders. Hanes *et al.* [21•] isolated picomolar affinity scFvs from a synthetic naïve library by combining the intrinsic selection and evolution power of ribosome display. An 'off-rate' selection procedure may further favor protein affinity maturation [36]. In off-rate selection, a pool of polypeptides is bound to an immobilized ligand. By adding an excess of free ligand, every dissociating library member will be immediately trapped. After incubation, only those binders with the lowest off-rate will remain bound to the immobilized antigen. Thus, increasing the incubation time with the competitive ligand increases the selection pressure applied. As the on-rate normally only changes within a relatively small window, lowering the off-rate will result in increased binding affinity. Using such off-rate selections over a period of up to ten days, Jermutus *et al.* [22••] were able to improve an antifluorescein scFv that already had a high initial affinity of 1.1 nM a further 30-fold. Using the same strategy, peptide-binding scFvs were evolved to affinities in the low picomolar range (C Zahnd *et al.*, unpublished data). These results demonstrate that off-rate selection is a valuable tool to select high-affinity binders from libraries. Interestingly, the mRNA — normally thought to be a very labile entity — was stable under these experimental conditions for more than 15 days.

Taken together, these results demonstrate that *in vitro* display technologies are not only valuable tools for the selection of binding molecules, but also for protein affinity maturation — either of a given molecule or in conjunction with the selection process from the initial library. This may have important implications in biotechnological and medical applications.

Maturation of protein stability

A common requirement for most biotechnological and medical applications of proteins is that they possess an intrinsic high stability against denaturation. Stability engineering is still a difficult task [37–40]. Evolutionary methods to perform stability engineering have shown promise, especially methods that employ phage display (reviewed in [41]). In a model system using antibody scFv fragments, Jermutus *et al.* [22••] have shown that ribosome display may be a valuable tool for *in vitro* evolution of protein stability. Antibody scFv fragments were evolved that are stable in the absence of disulfide bonds, which are normally required for their stability [37]. When the disulfide bond was allowed to reform, these scFv mutants were more stable than the corresponding wild-type protein, as indicated by urea denaturation experiments. They gave higher yields of functional protein upon periplasmic expression in *E. coli*, where disulfide bonds do form. Most importantly, the selected scFv mutants could also be functionally expressed in the reducing environment of the cytoplasm; an uncommon feature of an antibody scFv fragment. This study illustrates the versatility of the ribosome display approach — expression and selection can take place in a cytoplasmic-like environment, when dithiothreitol is added. Such stable and well-behaved antibodies might find application in tumor targeting [42] and as effective intrabodies [43,44] for the intracellular inactivation of proteins. Stability engineering of proteins by using heat or proteases as selection pressure (reviewed in [41]) may also be achievable with *in vitro* display technologies, especially for mRNA display.

Selection for enzymatic activity

It has been stated several times [6,41,45,46] that a combination of directed evolution and the use of display technologies provides a powerful strategy to evolve improved biocatalysts. Although it is known that enzymes can be functionally displayed on the ribosome [47], ribosome display had so far not been used to select for enzymatic activity. In this technique the genetic information (i.e. the mRNA) is not covalently attached to the protein. Thus, the mRNA can be simply eluted, even in applications based on suicide inhibitors, where the selected protein is covalently bound to the target. P Amstutz *et al.* (unpublished data) have performed a selection for enzymatic activity using ribosome display. Using a -lactamase suicide inhibitor, an active RTEM- -lactamase was successfully enriched over an inactive mutant. In these experiments the efficiency of activity selection was comparable to selection for affinity using a -lactamase ligand. Overall, *in vitro* display methods may open new roads for the selection of catalytically active proteins.

Display of cDNA products

Phage display and two-hybrid systems are well-established methods to screen or select cDNA libraries for binders [1,2,48,49]. Recently, two groups investigated the potential of *in vitro* display techniques for the display of cDNA products. Bieberich *et al.* [50] reported the specific isolation of the cDNA of sialyltransferase II by functional binding of the encoded enzyme to its substrate, ganglioside GD3, in a single-tube coupled ribosome display system. It remains unclear, however, if their ribosome display construct is free of a stop codon and if it contains an appropriate C-terminal tether. The demonstration that ribosome display can be performed in a single well of a microtiter plate may have implications for proteomic applications where automation and high-throughput screening are essential. By using cDNA product libraries displayed on mRNA, Hammond *et al.* [34••] isolated both previously known and several novel binders of the antiapoptotic protein Bcl-X_L. The binding affinities of these isolated proteins ranged from approximately 2 nM to 10 M.

In contrast to phage display or two-hybrid systems, *in vitro* display techniques are not biased by cytotoxic or secretion-incompatible cDNA products. In addition, *in vitro* display libraries can be preselected (see above) to improve their quality. Taken together, *in vitro* display of cDNA product libraries may be an interesting approach for proteomic applications, where the ultimate objectives are to functionally display all proteins and to minimize any selection or expression bias.

Conclusions

In vitro display technologies, namely ribosome and mRNA display, prove to be valuable tools for many applications other than merely selecting polypeptide binders. They have great potential for directed evolution of protein stability and affinity, the generation of high-quality libraries by *in vitro* preselection, the selection of enzymatic activities, and the display of cDNA and random-peptide libraries. In addition, these technologies have several features that should make them amenable to standardization and automation: they comprise fast selection cycles, allow the processing of huge libraries, are not limited by cellular transformations, and are not biased by *in vivo* environments. We envision that *in vitro* display technologies will have a great impact on applications in biotechnology, medicine and proteomics.

Acknowledgements

We thank Markus Kurz and Philip W Hammond for sharing unpublished results and Stephen F Marino and Christiane Schaffitzel for critically reading the manuscript and helpful suggestions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Mendelsohn AR, Brent R: Protein interaction methods — toward an endgame. *Science* 1999, 284:1948-1950.
2. Dunn IS: Phage display of proteins. *Curr Opin Biotechnol* 1996, 7:547-553.
3. Ghadessy FJ, Ong JL, Holliger P: Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA* 2001, 98:4552-4557.
4. Doi N, Yanagawa H: STABLE: protein–DNA fusion system for screening of combinatorial protein libraries ***in vitro***. *FEBS Lett* 1999, 457:227-230.
5. Tawfik DS, Griffiths AD: Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 1998, 16:652-656.
6. Griffiths AD, Tawfik DS: Man-made enzymes — from design to ***in vitro*** compartmentalisation. *Curr Opin Biotechnol* 2000, 11:338-353.
7. Plückthun A, Schaffitzel C, Hanes J, Jermutus L: ***In vitro*** selection and evolution of proteins. *Adv Protein Chem* 2000, 55:367-403.
8. Hanes J, Jermutus L, Plückthun A: Selecting and evolving functional proteins ***in vitro*** by ribosome display. *Methods Enzymol* 2000, 328:404-430.
9. Liu R, Barrick JE, Szostak JW, Roberts RW: Optimized synthesis of RNA–protein fusions for ***in vitro*** protein selection. *Methods Enzymol* 2000, 318:268-293.
10. Roberts RW: Totally ***in vitro*** protein selection using mRNA–protein fusions and ribosome display. *Curr Opin Chem Biol* 1999, 3:268-273.
11. Schaffitzel C, Hanes J, Jermutus L, Plückthun A: Ribosome display: an ***in vitro*** method for selection and evolution of antibodies from libraries. *J Immunol Methods* 1999, 231:119-135.
12. Jermutus L, Ryabova LA, Plückthun A: Recent advances in producing and selecting functional proteins by using cell-free translation. *Curr Opin Biotechnol* 1998, 9:534-548.
13. Mattheakis LC, Bhatt RR, Dower WJ: An ***in vitro*** polysome display system for identifying ligands from very large peptide libraries. *Proc Natl Acad Sci USA* 1994, 91:9022-9026.
14. Hanes J, Plückthun A: ***In vitro*** selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* 1997, 94:4937-4942.
15. He M, Taussig MJ: Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for ***in vitro*** display and evolution of antibody combining sites. *Nucleic Acids Res* 1997, 25:5132-5134.
16. Roberts RW, Szostak JW: RNA–peptide fusions for the ***in vitro*** selection of peptides and proteins. *Proc Natl Acad Sci USA* 1997, 94:12297-12302.
17. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H: ***In vitro*** virus: bonding of mRNA bearing puromycin at the 3 -terminal end to the C-terminal end of its encoded protein on the ribosome ***in vitro***. *FEBS Lett* 1997, 414:405-408.
18. Kurz M, Kuang G, Lohse PA: An efficient synthetic strategy for the preparation of nucleic acid-encoded peptide and protein libraries for ***in vitro*** evolution protocols. *Molecules* 2000, 5:1259-1264.
19. Kurz M, Kuang G, Lohse PA: Psoralen photo-crosslinked mRNA–puromycin conjugates: a novel template for the rapid and facile preparation of mRNA–protein fusions. *Nucleic Acids Res* 2000, 28:E83.
20. Hanes J, Jermutus L, Weber-Bornhauser S, Bosshard HR, Plückthun A: Ribosome display efficiently selects and evolves high-affinity antibodies ***in vitro*** from immune libraries. *Proc Natl Acad Sci USA* 1998, 95:14130-14135.
21. Hanes J, Schaffitzel C, Knappik A, Plückthun A: Picomolar affinity antibodies from a fully synthetic naïve library selected and evolved by ribosome display. *Nat Biotechnol* 2000, 18:1287-1292.
The authors selected a range of different scFvs with affinities up to 82 pM from a fully synthetic naïve antibody scFv library using ribosome display. All of the selected antibodies accumulated beneficial mutations throughout the selection cycles. This work demonstrated that ribosome display not only allows the selection of library members but also further evolves them, thereby mimicking the strategy of the immune system.
22. Jermutus L, Honegger A, Schwesinger F, Hanes J, Plückthun A: Tailoring ***in vitro*** evolution for protein affinity or stability. *Proc Natl Acad Sci USA* 2001, 98:75-80.
The authors demonstrate the potential of ribosome display for directed *in vitro* protein evolution. By combining ribosome display with DNA shuffling the authors improved an scFv 30-fold to a final affinity of 40 pM, using off-rate selections over a period of several days. In a second set of similar experiments they evolved an scFv to be functionally expressed under reducing conditions. Under these conditions the scFv evolved novel stabilizing structures to compensate for the loss of the disulfide bonds. The selected mutants, when allowed to reform disulfide bonds, showed improved stability (from an initial 24 kJ/mol to 54 kJ/mol).
23. Irving RA, Coia G, Roberts A, Nuttall SD, Hudson PJ: Ribosome display and affinity maturation: from antibodies to single V-domains and steps towards cancer therapeutics. *J Immunol Methods* 2001, 248:31-45.
24. He M, Menges M, Groves MA, Corps E, Liu H, Brüggemann M, Taussig MJ: Selection of a human anti-progesterone antibody fragment from a transgenic mouse library by ARM ribosome display. *J Immunol Methods* 1999, 231:105-117.
25. Kurz M, Gu K, Al-Gawari A, Lohse PA: cDNA–protein fusions: covalent protein–gene conjugates for the ***in vitro*** selection of peptides and proteins. *Chem Biochem* 2001, in press.
The authors describe a method to replace the mRNA in the mRNA–protein fusion with its cDNA. The cDNA–protein complex is significantly more stable than the corresponding mRNA complex.
26. Virnekäs B, Ge L, Plückthun A, Schneider KC, Wellnhofer G, Moroney SE: Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res* 1994, 22:5600-5607.

27. Cho G, Keefe AD, Liu R, Wilson DS, Szostak JW: Constructing high complexity synthetic libraries of long ORFs using ***in vitro*** selection. *J Mol Biol* 2000, 297:309-319.
This paper describes a strategy for improving protein libraries, by selecting against frame-shifts and internal stop codons. mRNA display was used to perform such preselection of three libraries: a random sequence, a patterned sequence and an $(\ /\ )_8$ (TIM) barrel library. Modules of these libraries were inserted between a C-terminal and N-terminal polypeptide tag. Selection for these tags yielded full-length in-frame protein modules, which were subsequently combined to form the libraries for selection. Using this strategy the proportion of correct full-length library members was increased by up to two orders of magnitude. The final library, originating from a 10 ml *in vitro* translation, had a complexity of around 10^{13} members.

28. Keefe AD, Szostak JW: Functional proteins from a random-sequence library. *Nature* 2001, 410:715-718.
From a library of 6×10^{12} polypeptides, consisting of a stretch of 80 completely random amino acids, four new ATP-binding folds were selected by mRNA display. By further mutagenesis and selection the behavior of these proteins was improved, yielding specific binders with affinities up to 100 nM. One selected fold seems to have a fold-stabilizing Zn^{2+}-binding site.

29. Wilson DS, Keefe AD, Szostak JW: The use of mRNA display to select high-affinity protein-binding peptides. *Proc Natl Acad Sci USA* 2001, 98:3750-3755.

30. Schmidt TG, Koepke J, Frank R, Skerra A: Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J Mol Biol* 1996, 255:753-766.

31. Zaccolo M, Gherardi E: The effect of high-frequency random mutagenesis on ***in vitro*** protein evolution: a study on TEM-1 -lactamase. *J Mol Biol* 1999, 285:775-783.

32. Cadwell RC, Joyce GF: Mutagenic PCR. *PCR Methods Appl* 1994, 3:S136-S140.

33. Minshull J, Stemmer WP: Protein evolution by molecular breeding. *Curr Opin Chem Biol* 1999, 3:284-290.

34. Hammond PW, Alpin J, Rise CE, Wright M, Kreider BL: ***In vitro*** selection and characterization of Bcl-X_L-binding proteins from a mix of tissue-specific mRNA display libraries. *J Biol Chem* 2001, 276:20898-20906.
After four rounds of mRNA display using uniquely tagged cDNA libraries from different tissues, 71 protein binders were selected against the anti-apoptotic protein Bcl-X_L. Of these, only eight were identified as false positives, as they were derived from introns or wrong reading frames. In addition to known binders of Bcl-X_L, several proteins not previously demonstrated to interact with Bcl-X_L were identified and their biological relevance can now be tested. This is the first successful report where novel binders of a target protein were selected from a cDNA product library by mRNA display, illustrating the potential of *in vitro* display technologies for proteomic applications.

35. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Plückthun A: ***In vitro*** generated antibodies specific for telomeric guanine-quadruplex DNA react with ***Stylonychia lemnae macronuclei***. *Proc Natl Acad Sci USA* 2001, in press.
This paper describes the ribosome display selection of high-affinity antibodies specific for guanine quadruplex DNA from a naïve library. Antibody scFv fragments recognizing different conformations of this DNA structure were selected and applied *in vivo* in ciliates. This work provides the first evidence for the occurrence of guanine quadruplex DNA in macronuclei of the ciliate *Stylonychia lemnae*.

36. Hawkins RE, Russell SJ, Winter G: Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J Mol Biol* 1992, 226:889-896.

37. Wörn A, Plückthun A: Stability engineering of antibody single-chain Fv fragments. *J Mol Biol* 2001, 305:989-1010.

38. Vieille C, Zeikus GJ: Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 2001, 65:1-43.

39. Lehmann M, Pasamontes L, Lassen SF, Wyss M: The consensus concept for thermostability engineering of proteins. *Biochim Biophys Acta* 2000, 1543:408-415.

40. Colacino F, Crichton RR: Enzyme thermostabilization: the state of the art. *Biotechnol Genet Eng Rev* 1997, 14:211-277.

41. Forrer P, Jung S, Plückthun A: Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Curr Opin Struct Biol* 1999, 9:514-520.

42. Houghton AN, Scheinberg DA: Monoclonal antibody therapies — a 'constant' threat to cancer. *Nat Med* 2000, 6:373-374.

43. Cattaneo A, Biocca S: The selection of intracellular antibodies. *Trends Biotechnol* 1999, 17:115-121.

44. Chames P, Baty D: Antibody engineering and its applications in tumor targeting and intracellular immunization. *FEMS Microbiol Lett* 2000, 189:1-8.

45. Petrounia IP, Arnold FH: Designed evolution of enzymatic properties. *Curr Opin Biotechnol* 2000, 11:325-330.

46. Olsen M, Iverson B, Georgiou G: High-throughput screening of enzyme libraries. *Curr Opin Biotechnol* 2000, 11:331-337.

47. Kolb VA, Makeyev EV, Spirin AS: Co-translational folding of a eukaryotic multidomain protein in a prokaryotic translation system. *J Biol Chem* 2000, 275:16597-16601.

48. Santi E, Capone S, Mennuni C, Lahm A, Tramontano A, Luzzago A, Nicosia A: Bacteriophage display of complex cDNA libraries: a new approach to functional genomics. *J Mol Biol* 2000, 296:497-508.

49. Fields S: Proteomics. Proteomics in genomeland. *Science* 2001, 291:1221-1224.

50. Bieberich E, Kapitonov D, Tencomnao T, Yu RK: Protein-ribosome-mRNA display: affinity isolation of enzyme-ribosome-mRNA complexes and cDNA cloning in a single-tube reaction. *Anal Biochem* 2000, 287:294-298.

Ribosome Display: *In vitro* Selection of Protein-Protein Interactions

Patrick Amstutz, Hans Kaspar Binz, Christian Zahnd and Andreas Plückthun*

I. Introduction

Ribosome display is an *in vitro* technology to identify and evolve proteins or peptides binding to a given target (Fig. 1) (Hanes et al., 2000a). While most selection technologies need living cells to achieve the essential coupling of genotype and phenotype, ribosome display uses the ribosomal complexes formed during *in vitro* translation to generate the physical coupling between polypeptide (phenotype) and mRNA (genotype)(Amstutz et al., 2001). Hence, no transformation step limiting the size of the usable library is necessary, allowing the selection from very large combinatorial libraries. In addition, the rapid selection cycles require an integral PCR step, which can be used for randomization, making this method ideal for directed evolution experiments. The fact that the ribosomal complex used for selection is not covalent allows an uncomplicated separation of the mRNA from the selected ribosomal complexes, even if the selected molecules bind the target with very high affinity or are even trapped covalently (Amstutz et al., 2002; Jermutus et al., 2001). All these benefits make ribosome display a good alternative to other selection techniques such as phage display (Smith, 1985).

Ribosome display has been successfully applied for the selection of peptides (Matsuura and Plückthun, 2003; Mattheakis et al., 1994), as well as folded proteins such as antibody fragments (Hanes and Plückthun, 1997; He and Taussig, 1997; Irving et al., 2001). Ribosome display can also be considered for the screening of cDNA libraries for interaction partners. Ribosome display ultimately selects always for a specific binding event. However, by carefully designing the selection pressure, molecules can be selected for many other parameters, such as enzymatic turnover (by selection with a suicide inhibitor, or active site ligand) (Amstutz et al., 2002; Takahashi et al., 2002), protein stability (by selecting for binding under conditions where most library members will not fold) (Jermutus et al., 2001), or protein biophysical properties (resistance to proteases and non-binding to hydrophobic surfaces) (Matsuura and Plückthun, 2003). It is the combination of this array of selection pressures with the convenient PCR-based randomisation techniques that makes ribosome display a powerful and versatile technology.

* Department of Biochemistry University of Zürich, Winterthurerstr. 190, CH 8057 Zürich, Switzerland
Andreas Plückthun (corresponding author), e-mail: plueckthun@bioc.unizh.ch, Tel (+41) 1 635 5570, Fax (+41) 1 635 5712
Patrick Amstutz, e-mail: pat7@access.unizh.ch, Tel. (+41) 1 635 55 76
Hans Kaspar Binz, e-mail: kbinz@bioc.unizh.ch, Tel. (+41) 1 635 55 76
Christian Zahnd , e-mail: zahnd@access.unizh.ch, Tel. (+41) 1 635 55 77

Keywords: Ribosome display, *In vitro* selection, Directed evolution, Cell-free translation, Combinatorial library

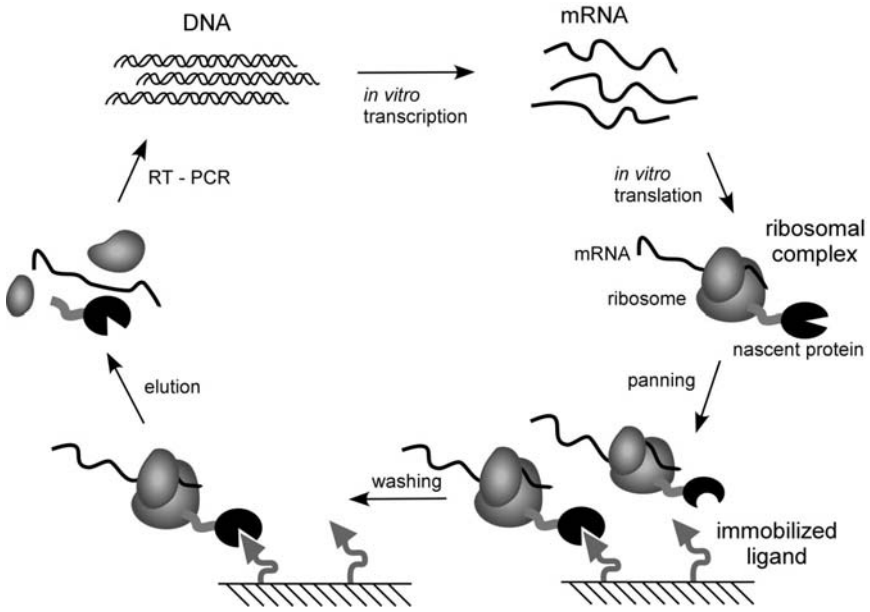


Figure 1. Ribosome display selection cycle. The DNA of the library of interest, fused in frame to a spacer carrying no stop codon, is transcribed *in vitro*. The resulting mRNA is used for *in vitro* translation. After a short time of translation (a few minutes) the ribosomes have probably run to the end of the mRNA and synthesized the encoded protein, but because of the absence of the stop codon, the protein remains connected to the tRNA. Stopping the translation reaction in ice-cold buffer with a high Mg^{2+} concentration stabilizes this ternary complex, consisting of mRNA, ribosome and nascent protein. The spacer, occupying the ribosomal tunnel, enables the domain of interest to fold on the ribosome. These ribosomal complexes are used for affinity selection. After washing, the mRNA of the selected complexes is released by complex dissociation. The genetic information of binders is rescued by RT-PCR, yielding a PCR product ready to go for the next selection cycle.

II. Materials and Instrumentation

A. Reagents

The following chemicals and enzymes are necessary to prepare the extract and to perform ribosome display selections: Luria broth base (GibcoBRL, 12795-084); agarose (Invitrogen 30391-023); glucose (Fluka 49150); potassium dihydrogen phosphate (KH_2PO_4, Fluka 60230); di-potassium hydrogen phosphate ($K_2HPO_4 \cdot 3H_2O$, Merck 1.05099.1000); yeast extract (GibcoBRL 30393-037); thiamine (Sigma T-4625); Tris-acetate (Serva 37190); magnesium acetate (MgAc, Sigma M-0631); potassium acetate (KAc, Fluka 60034); L-glutamic acid monopotassium salt monohydrate (KGlu, Fluka 49601) 20 natural amino acids (Sigma LAA-21 kit); adenosinetriphosphate (ATP, Roche Diagnostics 519 987); phosphoenolpyruvate trisodium salt (PEP, Fluka 79435); pyruvate kinase (Fluka 83328); GTP (Sigma, G-8877); cAMP (Sigma A-6885); acetylphosphate (Sigma A-0262); *E. coli* tRNA (Sigma R-4251); folinic acid (Fluka 47612); PEG 8000 (Fluka 81268); 1,4-dithio-treitol (DTT, Promega, V3155); sodium chloride (NaCl, Fluka, 71376); Tween-20 (Sigma, P-7949); neutravidin (Pierce 31000); bovine serum albumin (BSA, Fluka 05476), *Sacchoromyces cerevisiae* RNA (Fluka 83847); ribonuclease inhibitor RNasin (Promega N211B); reverse transcriptase Stratascript (50 U/ μl, Stratagene 600085-51); 10x Stratascript buffer (Stratagene 600085-52); dNTPs (5 mM of each dNTP, Eurogentec NU-0010-50); DNA-polymerase for PCR (e.g., Vent® polymerase, NEB M0254L); PCR buffer (e.g., thermopol buffer,

delivered with Vent® polymerase), dimethylsulfoxide (DMSO, Fluka 41640); NTPs (50 mM, Sigma); nitrocefin (Calbiochem 484400); Hepes (Sigma H-3375); spermidine (Sigma S-2501); T7 RNA polymerase (NEB M0251L); lithium chloride (LiCl, Fluka 62476); 100% ethanol (EtOH); sodium acetate (NaAc, Fluka 71180); heparin (Fluka 51550); disodium ethylenediaminetetraacetate (EDTA, Fluka 03680); T4 DNA-ligase (MBI Fermentas EL0011); 4-morpholinopropanesulfonic acid (MOPS, Fluka 69949); boric acid (Fluka 15660); guanidine thiocyanate (Fluka 50990); N,N-dimethyl formamide (Sigma Aldrich 27.054-7); 37% formaldehyde (Fluka 47629); UHP water; if proteins are displayed that depend on the correct formation of disulfide bonds protein disulfide isomerase should be used (PDI; Sigma P3818); ^{35}S-methionine (PerkinElmer NEG009H); triethylamine (Sigma-Aldrich 90335); OptiPhase2 scintillation liquid (PerkinElmer 1200-436)

B. Buffers

The following buffers are used in standard ribosome display selection rounds and we advise to prepare stocks: Tris-buffered saline (TBS: 50 mM TrisHCl pH 7.4 at 4 °C; 150 mM NaCl), TBS with Tween (TBST: TBS with 0.05 % (500 µl/l) Tween-20), washing buffer with Tween (WBT: 50 mM Tris-acetate pH 7.5 at 4°C; 150 mM NaCl; 50 mM MgAc; 0.05% Tween-20) and elution buffer (EB: 50 mM Tris-acetate pH 7.5 at 4°C; 150 mM NaCl; 50 mM EDTA); 10 x MOPS (0.2 M MOPS pH 7, 50 mM sodium acetate, 10 mM EDTA); 10 x TBE buffer: (89 mM Tris-buffered saline, 89 mM boric acid, 10 mM EDTA)

Note: the washing buffer used for ribosome display can be adjusted to any particular requirements given by the target molecule. Different buffer salts and detergents are compatible, only the Mg^{2+} concentration should be held at around 50 mM. If a new buffer composition is applied, a test ribosome display selection round with a known binder is recommended to determine compatibility.

C. Oligonucleotides

αssrA DNA (200 µM, 5'- TTAAGCTGCTAAAGCGTAGTTTTCGTCGTTTGCGACTA-3', standard quality)

T7B : Forward RD primer. Introduces T7 promotor and part of the 5'-loop (100 µM, 5'-ATACGAAATTAATACGACTCACTATAGGGAGACCACAACGG-3')

SDplus: Forward RD primer. Introduces the Shine-Dalgarno sequence and connects the T7 promoter with the FLAG-tag: 5'-AGACCACAACGGTTTCCCAATAATTTTGTTTAACTTTAAGAAGGAGATAT-ATCCATGGCGGACTACAAAGATGACG-3'

tolAk: Reverse primer for RD used with tolA as spacer introducing a stabilizing 3'-loop (100 µM 5'-CCGCACACCAGTAAGGTGTGCGGTTTCAGTTGCCGCTTTCTTTCT-3')

RDlinktolA: 5'-GGGGAAAGCTTTATATGGCCTCGGGGGCCGAATTCGAATCTGGTGGCCAGAAG-CAAGCTGAAGAGGCG-3'

Primers (reverse and forward) specific for the library of interest, which must introduce appropriate restriction sites for ligation into the ribosome display vector.

D. Bacterial Strain and Plasmid

We use *E. coli* strain MRE600 for the preparation of the extract. This strain is RNase I deficient (Kushner, 2002) and does not contain any antibiotic resistance (Wade and Robinson, 1966).

Ribosome display vector (pRDV), containing β-lactamase as insert (gene bank accession: AY327136).

E. Laboratory Equipment and Hardware

The following material will be used in ribosome display: ART filter pipet tips (10 µl, 20 µl, 200 µl, 1000 µl, nucleic acid and nuclease free tips, Molecular Bioproducts); QIAquick PCR purification and gel extraction kit (QIAgen 28104 and 28704); Maxisorp plate (Nunc-Immuno™ plate, Nunc 430341); Step pipet (Eppendorf Multipipette Plus 4981 000.019) with 5 ml and 10 ml tips (Eppendorf 0030 069.250 and 0030 069.269); Plastic seal (Carning Inc., Costar® 6524); RNase free 1.5 ml reaction vials (MolecularBioProducts 3445); Roche high pure RNA isolation kit (Roche 1 828 655); 0.2 mm syringe filter (Millipore SLGPR25KS); Dialysis tubing with a molecular weight cut-off of 6000-8000 Da. (for example: Spectrum Laboratories SpectraPor 132 650)

Furthermore, standard laboratory equipment is needed such as: Sorvall RC-5C Plus centrifuge with rotors SS-34 and GS-3 or equivalent; refrigerated table centrifuge; Shaker incubator; 5 l and 100 ml baffled shake flasks for *E. coli* culture; Emulsiflex (Avestin, Canada) or French Press (American Instrument Company, AMINCO); 4°C room; Liquid nitrogen (N_2); ELISA-plate shaker; UV/VIS spectrophotometer; Agarose gel electrophoresis system; Latex gloves, Speed-Vac (Savant Speed Vac Concentrator SVC100H); -20°C and –80°C freezer; Scintillation counter

III. Procedures

A. Preparation of S30-Extract

The preparation of the S30-Extract is performed according to Lesley, Zubay and Pratt, with minor modifications (Chen and Zubay, 1983; Lesley, 1995; Pratt, 1984; Zubay, 1973).

General consideration: 1 liter *E. coli* culture yields approximately 8 ml extract. If you plan to do ribosome display at a large scale, grow several cultures in parallel. It is important that the cells used for extract preparation are harvested in an early logarithmic phase. If the libraries used for selection contain disulfide bonds, one should omit DTT from the extract. If no disulfides need to be formed 1 mM DTT can be added to the S30 buffer as it increases the translation efficiency slightly.

Material, solutions and strain:

Solutions, Strain and Hardware:

E. coli strain MRE600 (Wade and Robinson, 1966); Luria broth base; Incomplete rich medium: 5.6 g/l KH_2PO_4, 37.8 g/l $K_2HPO_4 \cdot 3H_2O$, 10 g/l yeast extract, 15 mg/l thiamine - after autoclaving add 50 ml 40% (w/v) glucose sterile filtered; 0.1 M MgAc; 10x S30 buffer: 100 mM Tris-acetate pH 7.5 at 4 °C, 140 mM MgAc, 600 mM KAc - store at 4 °C or chill buffer in ice bath before use; 10 ml preincubation mix - must be prepared immediately before use: 3.75 ml 2 M Tris-acetate pH 7.5 at 4 °C, 71 µl 3 M

MgAc, 75 µl amino acid mix (10 mM of each of the 20 natural amino acids), 0.3 ml 0.2 M ATP, 0.2 g PEP, 50 U pyruvate kinase.
Material: 5 l baffled flasks; shaker at 37°C for *E. coli* culture; refrigerated centrifuges (GS-3, SS-34); dialysis tubing MW cut-off 6000 - 8000 Da.; Emulsiflex or French press.

Steps

Day 1:

1. Prepare an LB/glucose plate and streak out MRE600 on the plate. Grow it overnight at 37°C
2. Prepare all chemicals, media and buffers for *E. coli* extract preparation: Autoclave 1 l incomplete rich medium, 500 ml of LB/glucose medium and one 100 ml and one 5 l shake flask. Prepare 50 ml 40% glucose, 10 ml 0.1 M MgAc and 1 l 10x S30 buffer (use it as 1x S30 buffer afterwards). All buffers should be stored at 4°C.

Day 2:

3. Prepare an overnight preculture by inoculating 50 ml LB/glucose medium with a colony of MRE 600, which is shaken overnight at 37°C.

Day 3:

4. Add 1 l incomplete rich medium into the 5 l shake flask and add 50 ml 40% glucose and 10 ml 0.1M MgAc both by sterile filtration (0.2 µm syringe filter).
5. Inoculate the culture with 10 ml overnight culture (approximately 1%) and let it shake at 37°C to OD_{600} = 1.0 – 1.2. Then transfer the culture in an ice-water bath and quickly add 100 g ice (small shovel) to the culture. Shake the culture in the ice-water bath by hand for 5 minutes. Collect the cell pellet by centrifugation at 4°C (15 min, GS-3, 5000 rpm). Wash the pellet at least 3 times with 50-100 ml of S30 buffer.
6. Determine the weight of the pellets (typically 1-1.5 g/l). The cell pellet can now be shock-frozen in liquid nitrogen and stored at -80 °C until further processing. Do not store the pellets longer than 2 days.

Day 4:

Do wear gloves during all following steps!

7. Thaw the pellets on ice and resuspend the cells in S30 buffer (50 ml). Centrifuge at 4°C, full speed in an appropriate centrifuge to collect the cell pellet. Resuspend the cells in 4 ml/g (wet cell weight) S30 buffer.
8. Lyse the cells with one passage through an EmulsiFlex (approximately 17000 psi) or a French press (at 1000 psi). Repeating passages will decrease translation activity. Centrifuge the lysate at 4°C for 30 min (SS-34, 20000 g). Take the supernatant and repeat this centrifugation step.
9. Transfer the supernatant to a 50 ml Falcon tube. Add preincubation mix (1 ml/6.5 ml supernatant) and incubate at RT for 60 min by slowly shaking the tube. In this step most endogenous RNA and DNA will be degraded by nucleases and translation will run out.

10. Transfer the extract into a dialysis device (MW cut-off 6000-8000 Da) and dialyze the extract 3 times for at least 4 h each at 4°C against S30 buffer (500 ml). Dialyse overnight in a fourth step.

Day 5:

11. Transfer the extract into a single 50 ml tube. Aliquot the extract into RNase-free tubes immediately (for example 110 µl and 55 µl aliquots, use filter tips to avoid RNase contamination) and directly freeze the aliquots in liquid nitrogen. Store the aliquoted extract at -80 °C (it will be fully active for months to years).

B. Premix Preparation and Extract Optimization

The premix provides the S30-extract with all amino acids (except for methionine, see below), tRNAs, the energy regeneration system and salts, which are needed for translation. For optimal translation efficiency every premix should be adjusted to fit the corresponding extract, especially with respect to the concentration of magnesium, potassium, PEG8000 and amount of extract. The PremixA recipe given below contains only minimal concentrations of KGlu, MgAc and PEG-8000. By performing translations (compare section F) using a test mRNA and by gradually adding increasing amounts of these components, the translation efficiency of the extract will be optimized to its maximal activity. The optimization of the premix to the S30-extract is optimally done by translating the mRNA encoding an enzyme, whose activity is easily determined, such as e.g. β-lactamase. We routinely use a cysteine-free version of the enzyme (Laminet and Plückthun, 1989) in a ribosome display suitable format (described in section C) for optimization.

Solutions

You will need approximately equivalent amounts of premix and extract. PremixA: 250 mM Tris-acetate pH 7.5 at 4°C, 1.75 mM of each amino acid except methionine, 10 mM ATP, 2.5 mM GTP, 5 mM cAMP, 150 mM acetylphosphate, 2.5 mg/ml *E. coli* tRNA, 0.1 mg/ml folinic acid, 4 µM α–ssrA DNA. KGlu (180 - 220 mM), MgAc (10 - 15 mM) and PEG-8000 (5 - 15 % w/v) have to be adjusted to the corresponding extract.

β-lactamase assay buffer: Dissolve 5.3 mg nitrocefin in 250 µl DMSO and add this to 50 ml 50 mM potassium phosphate buffer (pH 7) (Laminet and Plückthun, 1989).

Steps

To avoid RNase contamination use filter tips and wear gloves for all of the following steps.

1. Mix all components to yield PremixA.
2. Incubate the premix in a waterbath at 37 °C to solubilize all components.
3. Perform different *in vitro* translations in parallel of β-lactamase mRNA as described below (Section F). Add increasing concentrations of MgAc (10 - 15 mM), KGlu (180 - 220 mM), PEG-8000 (5 - 15% w/v of premix) and amount of extract (30 - 50 µl for a 110 µl translation reaction).

4. First optimize the MgAc concentration, then the KGlu concentration and last the PEG-8000 concentration. The translation time relevant for optimization should be around 10 min. Stop translation by diluting the reaction 5 times in WBT.
5. To detect β-lactamase activity, use a nitrocefin assay. Use 10 - 20 µl of the stopped translation per ml β-lactamase assay buffer and follow the reaction with a photospectrometer at 486 nm.
6. After determining the conditions giving the highest activity, add the chemicals at the optimal concentration to the PremixA stock yielding PremixZ, optimized for this very batch of extract.
7. Aliquot the PremixZ in RNase-free tubes and shock freeze the samples in liquid nitrogen (for example 500 µl aliquots).

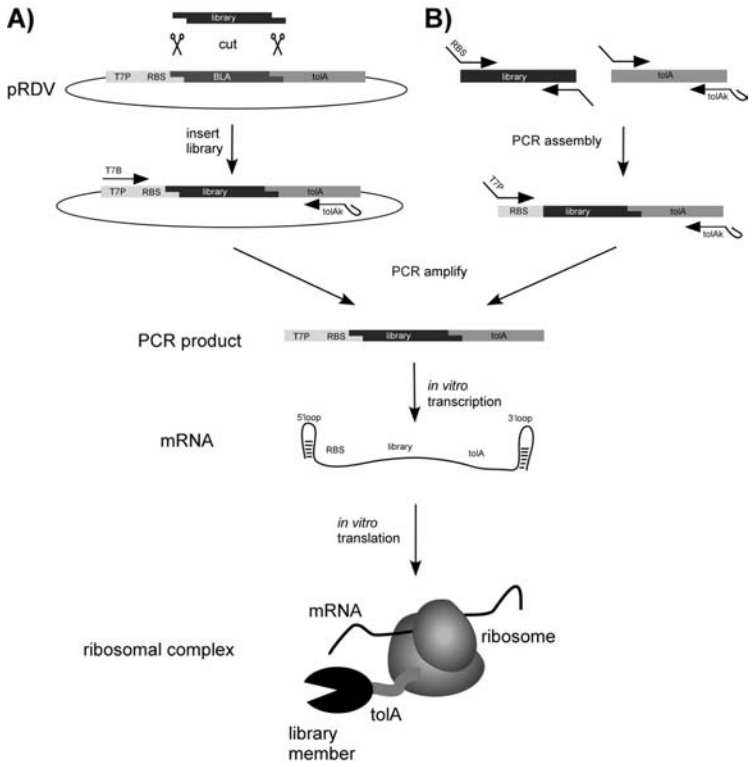


Figure 2. Generation of the ribosome display construct. For ribosome display the library of interest has to be flanked by an upstream promoter region and a C-terminal spacer carrying no stop codon (Fig. 3). **A)** The library is PCR amplified with primers carrying restriction sites suitable for ligation into the ribosome display vector (pRDV), which carries the necessary library flanking regions. The PCR product of the library is digested and ligated into pRDV. A second PCR on this ligation reaction with the primers T7B and tolAk yields a PCR product ready for *in vitro* transcription. **B)** Alternatively, the ribosome display construct can be generated by assembly PCR. The library and the spacer are PCR amplified separately with primers, so that the C-terminal part of the library and the N-terminal part of the spacer share overlapping sequences. An assembly PCR with the library and the spacer DNA, using appropriate primers, finally yields the ribosome display construct. *In vitro* transcription of the PCR product of either **A)** or **B)** yields mRNA carrying 5'- and 3'-stemloops (which make the mRNA more stable toward exonuclease digestion), a ribosome binding site (RBS), the library of interest and a spacer carrying no stop codon. By stopping the *in vitro* translation in ice-cold buffer with high Mg^{2+} concentration, stable complexes of mRNA, ribosome and nascent protein are formed, ready for panning.

C. Preparation of the Ribosome Display Construct

To perform ribosome display one needs a high quality library in the appropriate format. In this section we will not explain how to generate this library, as this entirely depends on the experimental goal, but how to convert an existing one into a format suitable for ribosome display.

A ribosome display construct is composed of a T7 promoter, followed by a ribosomal binding site and an open reading frame, which in turn consists of the library fused in frame to a C-terminal spacer polypeptide which has no stop codon (Fig. 2; Fig. 3). The lack of the stop codon prevents the binding of the termination factors TF-1, TF-2 and TF-3. A high magnesium concentration "sinters" the ribosome, which consists largely of folded RNA with a protein coat. The low temperature presumably prevents the hydrolysis of the peptidyl-tRNA and minimizes mRNA degradation. All these measures together ensure that the ternary complex of mRNA (genotype), ribosome and displayed protein (phenotype) remains stable. The C-terminal spacer (usually derived from tonB, tolA, M13 gpIII or pD), which will partially remain in the ribosomal tunnel, ensures that the library protein can fold and is displayed on the ribosome. The T7-promoter allows efficient *in vitro* transcription of the construct. A 5'- and a 3'-stemloop protect the mRNA against exonucleases.

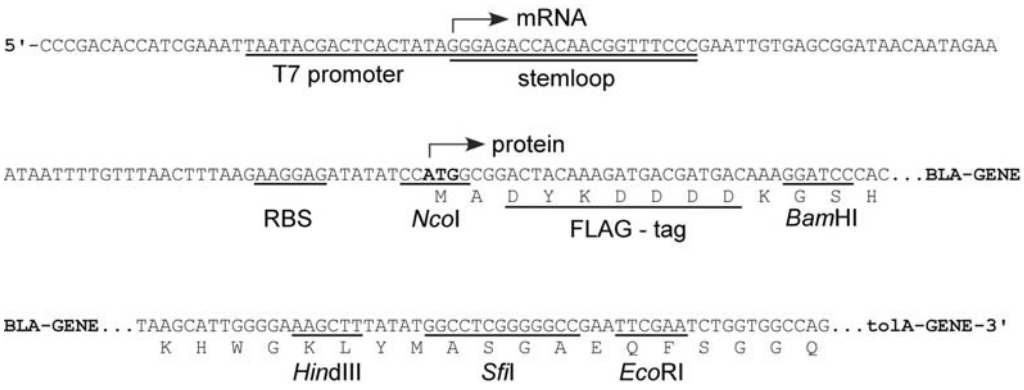


Figure 3. DNA sequence of the expression cassette of the ribosome display vector (pRDV). The mRNA is produced from a T7 promoter starting with a 5'-stemloop, with no additional overhang. The ribosome binding site (RBS), also called Shine-Dalgarno sequence, is located upstream of the start codon. The open reading frame consists of a FLAG-tag, the β-lactamase gene (serving as a dummy insert) in frame with a protein spacer, here *tolA*. Different restriction sites allow the cloning of the library into pRDV to replace the β-lactamase gene.

Solutions, Plasmids and Strains:

QIAquick PCR purification and gel extraction kit; pRDV; appropriate restriction enzymes; T4-DNA-ligase

a) Generation of the Ribosome Display Construct via pRDV

To accelerate the procedure of bringing a library into the ribosome display format, we generated a vector containing the necessary flanking regions (ribosome display vector, pRDV; Fig. 2). The library

is PCR amplified, cut with the appropriate restriction enzymes and ligated into the vector such that it is in frame with the spacer (Fig. 3). A second PCR on this ligation product directly amplifies the library with all features necessary for ribosome display: the T7-promoter, the RBS and the spacer without stop codon (Fig. 2). This PCR product is directly used for *in vitro* transcription, to yield the library-mRNA ready to go. The main advantages of the ribosome display vector are that one can generate large amounts of it (mini to maxi prep), it is easy to handle and always provides error-free library flanking regions. The use of the vector is not only interesting for the initial generation of the ribosome display construct, but also for the first selection rounds. If one only amplifies the library gene of the selected clones after the panning procedure without all the flanking regions, one is able to even recover library members partly degraded by RNases in the flanking regions. The recovered genes are then re-ligated into pRDV and one is again ready to go for another round of selection.

Steps:

1. Amplify the library with primers carrying the appropriate restriction sites for cloning into pRDV (Fig 3).
2. Digest the PCR product and purify it using QIAquick columns (≥150 ng; amount depending on the library size). pRDV is digested, optionally dephosphorylated and the backbone is agarose-gel purified. The vector insert has to be removed, as it would also ligate to the library DNA, decreasing the final complexity.
3. Ligate the PCR product of the library into pRDV (molar ratio insert:vector = 7:1).
4. Perform a PCR reaction on the ligation mix with the forward primer T7B (annealing on the T7 promoter with a stabilizing 5'-loop) and the reverse primer tolAk (annealing on the C-terminal spacer carrying a stabilizing 3'-loop).
5. Analyze the DNA on an agarose gel, checking size, purity and amount. If the band is sharp and indicates a concentration higher than 40 ng/µl, the PCR product is directly (i.e. without purification) used for *in vitro* transcription.

b) Generation of the Ribosome Display Construct via Assembly PCR

In some cases, it may be preferable not to ligate the library into pRDV, but to use PCR assembly to generate the ribosome display construct (Fig. 2). In this case, both the library and the spacer (e.g. tolA) are PCR amplified, so that the 3'-end of the library and the 5'-end of the spacer share overlapping sequences (Fig. 2). The PCR products of the library and the spacer are assembled and amplified with the primers SDplus (introducing the ribosome binding site and the connection to the T7-promoter) and tolAk. A final PCR reaction with the primers T7B (introducing the T7-promoter) and tolAk completes the construct, ready for transcription (Fig. 2).

Steps:

1. Amplify the library of interest with appropriate primers, introducing the FLAG-tag at the 5'-end of the library, on which the primer SDplus can anneal in step 3. The reverse primer should anneal on the library and add an overlap corresponding to RDlinktolA. There is no need to introduce restriction sites, however, we strongly recommend the use of the same primers as for the cloning

into the RDV, to have the possibility of using pRDV and to reduce the number of different fragments.

2. Amplify the ribosome display spacer using the primers RDlinktolA (forward) and tolAk (reverse). Thereby, the forward primer generates an overlap between the library and the spacer.
3. In an assembly-PCR reaction, the spacer is fused to the library. Mix DNA of the library with an excess of spacer DNA and perform 7 cycles of PCR without the addition of primers. For this step, we use a lower annealing temperature as for the normal amplification. After 7 cycles, add primers tolAk and SDplus to the reaction and perform another 25 cycles of amplification.
4. Isolate the full-length band from an agarose gel and amplify the band using the primers T7B and tolAk.
5. Analyze the DNA on an agarose gel, checking size, purity and amount. If the band is sharp and indicates a concentration higher than 40 ng/µl, the PCR product is directly (i.e. without purification) used for *in vitro* transcription.

D. Transcription of PCR Products

Solutions and Hardware:

5xT7 polymerase buffer: 200 mM Hepes pH 7.6, 30 mM MgAc, 2 mM spermidine, 40 mM DTT; NTPs (50 mM each); T7 RNA polymerase; RNasin; 6M LiCl; 70% EtOH; 100% EtOH; 3M NaAc; agarose; guanidine thiocyanate; formamide; 37% formaldehyde; MOPS;
Speed Vac; heating blocks; UV/VIS spectrophotometer

Steps

To avoid RNase contamination use filter tips and wear gloves for all of the following steps.

1. Mix the components for the following reaction using the PCR product from section C. The PCR product should not be purified. Add 20.0 µl 5x T7 polymerase buffer, 14.0 µl NTPs (50 mM each), 4.00 µl T7 RNA polymerase, 2.00 µl RNasin, 22.5 µl PCR product, 37.5 µl UHP. Let the transcription reaction run for 2-3 hours at 37 - 38°C.
2. Add 100 µl UHP and 200 µl 6 M LiCl, both ice-cold, and place on ice for 30 min, before centrifugation at 20000 g (4°C, 30 min). Discard supernatant and wash the pellet with 500 µl ice-cold 70% EtOH. Shortly dry the pellet on your bench (5 min, open lid) and take it up in 200 µl ice-cold UHP. Make sure it dissolved completely, before centrifuging at 20000 g (4°C, 5 min).
3. Transfer 180 µl supernatant to a new tube and add 20 µl 3 M sodium acetate and 500 µl ice-cold 100% EtOH. Keep the solution on ice or at -20°C for 30 min, before centrifuging it at 20000 g (4°C, 30 min). Discard the supernatant wash the pellet with 500 µl ice-cold 70% EtOH. Discard the supernatant and finally dry the pellet in a Speed-Vac apparatus.
4. Take up the mRNA pellet in 30 µl ice-cold UHP and make sure it dissolved completely. Take 2 µl of this solution and dilute to 500 µl with ice-cold UHP for OD_{260} quantification and immediately N_2-freeze the rest of the RNA for further use.

5. Immediately (the RNA will be degraded and the signal will increase) measure the OD_{260}. For RNA, an OD_{260} of 1 corresponds to a concentration of 40 μg/ml.
6. Add UHP to the RNA stock in order to reach a standard concentration of 2.5 μg/μl.
7. The RNA quality can optionally be checked by agarose gel electrophoresis. Cast an agarose gel (1.5 %, depending on your RNA size) adding 2% w/v of 1 M guanidine thiocyanate . Denature 5 μg (i.e. 2 μl) RNA for 10 min at 70°C in 15.5 μl loading buffer (10 μl formamide, 3.5 μl 37% formaldehyde, 2 μl MOPS), place it on ice and load the gel.

E. Target Molecule Immobilization

To perform a selection, the target molecule must be immobilized in a conformation relevant for further applications. A very promising way to do so is to biotinylate the target molecule and immobilize it via neutravidin or streptavidin. Biotinylation can be done chemically with commercially available reagents, either attaching the biotin to cysteine or lysine residues. The problem of this unspecific approach is that biotinylation might destroy epitopes. Alternatively, the target protein can be expressed in recombinant form with a biotinylation tag, i.e. a peptide sequence, which is recognized and biotinylated by the *E. coli* biotinylation enzyme BirA (Schatz, 1993). If for any reason biotinylation is no option, one can either immobilize the target molecule directly on the hydrophobic surface of a microtiter plate well, or use a specific antibody, which itself can easily be immobilized via protein A or G. Note that the buffers used here may have to be adapted to the needs of the target molecules.

Solutions and Hardware

TBS, TBST, WBT at 4°C

Maxisorp plate; plastic seal; step pipet; plate shaker; 20 μM stock neutravidin; BSA; target molecule of choice

Steps

To avoid RNase contamination use filter tips and wear gloves for all of the following steps. We also recommend carrying out the selection, RT and PCR in duplicate, to check the reproducibility of the selection. Therefore, one target molecule is routinely immobilized in two wells.

Day 1:

1. Wash a Maxisorp plate 3 times with TBS and beat dry. Pipet (with a step pipet) 100 μl of a 66 nM neutravidin solution in TBS into the wells. Seal with plastic and store overnight at 4°C. If the target molecule is not biotinylated it can also be directly immobilized as the neutravidin is, but it might denature at least partly during this procedure.

Day 2 (Day of the Ribosome Display Round):

2. Wash the plate incubated overnight with cold TBS three times. Add 300 μl 0.5% BSA in TBS to the wells (with a step pipet) to block all hydrophobic surfaces with BSA. Seal the plate with plastic and

incubate on a shaker for 1 h at RT. If you directly immobilized the target molecule, incubate the BSA solution 2 h at 4°C and proceed directly to step 5.

3. Wash the plate three times with TBS, beat dry and add 200 µl 0.5% BSA in TBS to each well (step pipet).
4. Add 5 µl of the target molecule (~10 µM) to the respective well. Use biotinylated target molecule void of free biotin. Seal with plastic and incubate on shaker for 1 h at 4°C. To avoid binders to BSA or neutravidin, it is recommended to immobilize the target molecule only in every second well, and use the alternate wells for pre-panning.
5. Wash the plate 4 times with TBS, to get rid of the unbound target molecule. Wash at least once with WBT (step pipet) to equilibrate the well with the ribosome display buffer. Incubate wells that will not directly be used for panning with WBT (amount of liquid as will be used for panning). Keep the plate at 4°C.

F. *In vitro* Translation

Solutions

UHP; WBT; WBT with 0.5% BSA; heparin (200 mg/ml); methionine (200 mM); S30 extract; PremixZ; the mRNA of the library; optionally PDI (4 mg/ml, reconstituted from lyophilized protein); all buffers should be kept on ice, the RNA in liquid nitrogen unless stated otherwise.

Steps

We recommend carrying out the panning in duplicate, to check the reproducibility of the selection. To avoid RNase contamination use filter tips and wear gloves for all of the following steps.

1. Mix the following components (amounts given are for one reaction): 13.0 µl UHP water, 2.0 µl Met (200 mM), 41.0 µl PremixZ (thaw on ice, vortex before pipetting) and 50.0 µl extract. Volumes might vary, depending on the batch of S30 extract. This mix can be kept on ice for a short period (few minutes). If disulfide bridges are to be formed: add 0.625 µl PDI (4 mg/ml).
2. Add 4 µl mRNA (2.5 µg/µl) into a fresh RNase-free tube and freeze it in liquid nitrogen.
3. Add the translation mix to the frozen mRNA, dissolve the pellet by flicking the tube and translate for 6-12 min at 37°C.
4. In the meantime, prepare 440 µl stopping buffer (WBT with 0.5% BSA and 12.5 µl/ml heparin (200 mg/ml)) in RNase-free tubes and put them on ice.
5. After the 6-12 min translation, pipet 100 µl translation into the ice-cold stopping buffer to stop translation (always keep the ribosomal complexes on ice or at 4°C).
6. Centrifuge at 20000 g, 4°C for 5 minutes and use the supernatant containing the ribosomal complexes for panning.

G. Panning

Solutions and Hardware

WBT; EB; yeast RNA (25 μg/μl); cold room; microtiter plate shaker; Roche High Pure RNA isolation kit.

Steps

All the following steps should be performed in a cold room (at 4°C or slightly below). The low temperature guarantees complex and mRNA stability. Under such conditions ribosomal complexes have survived off-rate selection procedures of 10 days and longer. During all binding and elution steps shake the microtiter plate gently. To avoid RNase contamination use filter tips and wear gloves.

1. Add 100 - 200 μl stopped translation mix to the well, where no target molecule is immobilized and incubate for 1 h. In this prepanning step all BSA-binding, neutravidin-binding or simply sticky complexes are removed. If unspecific complexes are causing problems, more than one pre-panning step should be done.
2. Transfer the solution to the well with the immobilized target molecule and incubate for 1 h.
3. Wash the well to remove non-binding complexes. The time span of the washing step and the number of washing steps equals the selection pressure. One usually starts with short washing times in the first rounds (only rinsing six times) and increases the washing periods for later rounds (up to 3-4 h). As washing is always a dilution, it is important to fill the wells to the top and exchange the buffer many times (at least six times, in all rounds!). If harsher selection pressure than simple washing is required, for example to select for affinities below nM, one should consider immobilisation of low target molecule concentrations, competitive elution (with free target molecule in the solution) or off-rate selection procedures (see section J).
4. Before the mRNA of the selected complexes is eluted, prepare 3 RNase-free tubes for each well to be eluted. One tube with 400 μl lysis buffer of the Roche RNA-purification kit to purify the RNA, one tube to collect the purified mRNA and one tube for reverse transcription. Also prepare one Roche-RNA purification column for each well to be eluted. Label all tubes and columns appropriately.
5. Calculate and prepare the amount of elution buffer you need (200 μl/well) and add 50 μg/ml *S. cerevisiae* RNA (2 μl of a 25 μg/μl stock per ml) to it. Keep the buffer on ice.
6. Elution is done in the cold room. Wash your well one last time with WBT, then remove the supernatant completely and beat the plate dry. Add 100 μl ribosome display elution buffer, shake for 10 minutes, then transfer this eluate to the tube containing the lysis buffer and mix well. Repeat this procedure with another 100 μl of elution buffer. In the lysis buffer, the RNA is stable and can be brought to room temperature.

H. RT-PCR

Solutions and Hardware

Roche High Pure RNA isolation kit; 100 mM DTT; RNasin; Stratascript (50 U/ µl; Stratagene); dNTPs; DNA-polymerase; oligonucleotides; 10x Stratascript buffer; polymerase buffer; DMSO; agarose.

Steps

The RNA purification is done according to the Roche protocol, with slight modifications. All the centrifugation steps are carried out a 4°C.

mRNA Purification

1. Just before the RNA-purification, thaw the reagents for the reverse transcription (DTT, 10x Superscript buffer, dNTPs, oligonucleotide for reverse transcription).
2. Set two heating blocks to 70°C and 50°C respectively.
3. Apply the lysis buffer/eluate mixture on the column and spin for 1 minute at 8000 g. Discard the flow-through and wash with 500 µl buffer 1 (black capped bottle in the Roche kit; the DNAse incubation step, described in the Roche protocol is not necessary). Discard the flow-through and wash with 500 µl buffer 2 (blue capped bottle in the Roche kit). Discard the flow-through. Add 100 µl buffer 2 (blue cap) and spin for 2 minutes at 13000 g. Transfer the column to a tube to collect the RNA.
4. Elute with 30 µl Roche elution buffer (at 8000 g for 1 min) and directly put the RNA containing collection tube to 70°C for 10 minutes to denaturate the RNA. During this incubation time, pipet the reverse transcription (i.e. directly proceed with step 5).

Reverse Transcription

5. In a mastermix tube, add the following components (amounts given are for one reaction): 0.25 µl reverse primer (100 µM), 0.50 µl dNTP (5 mM each), 0.50 µl RNasin, 0.50 µl Stratascript, 2.00 µl 10x Strascript buffer, 2.00 µl DTT (100 mM), 2.00 µl UHP. Distribute the RT-mix (7.75 µl) to the previously labelled tubes and keep them on ice.
6. Spin the denatured eluted RNA-samples shortly and set them on ice. Add 12.25 µl of the eluted RNA to the RT-mix (N_2-freeze the rest of the RNA directly after adding it to the reaction).
7. Place the RT-reaction on the 50°C heat block for 45 minutes.

PCR

8. Mix the following components (amounts given are for one reaction): 5.00 µl Thermopol buffer, 2.00 µl dNTPs, 2.50 µl DMSO, 1.00 µl forward primer (100 µM), 1.00 µl reverse primer (100 µM), 5.00 µl RT-template, 0.50 µl polymerase, 33.0 µl UHP. PCR results can be improved if a hot start is performed (adding the polymerase in the 5 min preincubation step at 95°C).

9. Run the following PCR program (to be adapted according to primers and template): 5 min at 95°C, X times (30 sec at 95°C, 30 sec at 50°C, 1 min at 72°C), 5 min at 72°C, 4°C infinitely. Adjust the number of cycles (X = 25 - 45) to the corresponding selection round.
10. Verify the PCR product quality by an appropriate agarose gel electrophoresis.
11. Usually the quality of the PCR product is not good enough to be used directly for *in vitro* transcription to generate mRNA for the next round. The desired band is gel purified and a second PCR is performed on the purified first PCR to yield high quality DNA.
12. If you wish to check the pool of selected clones for binders, ligate the selected library members after PCR amplification into a vector suitable for expression, transform *E. coli* to obtain individual clones and test binding specificity, for example by crude extract ELISA. To asses the specificity of the binders, one can either compare binding of the selected molecules to the specific target molecule to the binding to an unrelated molecule, or test the ability that the specific binding can be inhibited by adding the purified unbiotinylated target molecule. The level of inhibition gives a first crude estimate of the affinity.

Troubleshooting:

RT-PCR

If the RT-PCR with the forward (e.g. T7B) and reverse (e.g. tolAk) primers does not yield a high quality product, one can amplify only the coding region of the selected library members. This usually improves the yield and the quality of the PCR product. This is most likely due to the fact that RNAses degrade the mRNA from the ends. Amplifying only the central (library) stretch can rescue partly degraded clones. The PCR product of the library stretch is subsequently religated into pRDV as described in section C. As this procedure rescues more clones than the PCR of the whole construct, we would recommend it for the first rounds, to guarantee that no binders are lost.

Panning Controls

Ribosome display has many error-sensitive steps. It is therefore recommended to do the panning, RT and PCR in duplicate, to check the reproducibility of the selection. Specific binders should be enriched from round to round. This should correlate with the number of PCR cycles needed to amplify the DNA after RT, which should decrease from round to round. Usually one can reduce the cycle number by around five per round. If the selection pressure is increased, the yield will drop. When enrichment is observed (for ankyrin repeat protein libraries, e.g. after round two, while for antibody libraries after round 3 to 4), one can test the specificity of the selected pool by comparing panning results against the correct and an unrelated target molecule. If the pool is specific, only the correct target molecule will give a PCR product after RT. If the pool gives also a signal with the unspecific target molecule, the majority of the clones are still unspecific. However, there might also be a population of specific binders in the pool. An additional panning round with increased pre-panning can reduce the background. Alternatively, single clone analysis might directly yield specific binders.

I. Radioimmunoassay (RIA)

The radioimmunoassay (RIA) is another fast and convenient method to check whether specific target binding molecules have been enriched in a pool (Hanes et al., 1998). It can be used for the evaluation of both selected pools and individual binders. The RNA of a pool or single binder is translated in the presence of radioactively labelled ^{35}S-methionine. Therefore, the radioactive protein which binds to the surface-immobilized target molecule can easily be quantified. The binding should be performed in the presence and absence of soluble competitor target molecules, and a control for unspecific binding should be included. In the competition assay, the minimal concentration of competitor still leading to half inhibition of the maximal binding signal is a crude measure for the affinity. Therefore, RIA facilitates the ranking of the affinities of different clones isolated after affinity maturation.

Solutions and Hardware

The general handling is the same as for the *in vitro* translation. However, the radioactive material must be handled with the appropriate precautions. Do the radioactive work in a designated area of your laboratory. We recommend to use filter tips and to wear gloves during the experiment. We recommend doing the RIA in duplicate.

UHP; WBT; WBT with 0.5% BSA; milk powder; heparin (200 mg/ml); ^{35}S-methionine (10 mCi/ml, 1175 Ci/mmol); S30 extract; PremixZ; mRNA of a selected pool or single binders; pRDV; optionally PDI (4 mg/ml); 0.1 M triethylamine; liquid scintillation cocktail “OptiPhase2”.

Scintillation counter, ELISA-plate shaker

Steps

RNA Preparation

1a. The analysis of whole pools of potential binders can be performed similar to a normal ribosome display round. A PCR reaction is performed on the selected pool but with primers introducing a 3’-stop codon and the standard ribosome display 5’-end (including T7-promotor and ribosome binding site). From this PCR product, RNA is produced as described in section D.

1b. Alternatively, if you wish to analyze single binders, the PCR product from step 1a is digested with the appropriate restriction enzymes and ligated into the ribosome display vector (RDV). After transformation, plasmids are isolated from single colonies. The transcription can be performed directly from the plasmid, which should be present at a concentration of at least 100 ng/µl.

RIA

2. Prepare a Maxisorp plate with immobilized target molecules as described in section E.
3. Perform *in vitro* translations as described in section F with the following modifications: instead of cold methionine use 2 µl of ^{35}S-methionine (0.3 µM, 50 µCi/ml final concentration) per 110 µl reaction and perform the translation for 30-45 min at 37°C.
4. Stop the translation with 440 µl WBT and centrifuge the samples for 5 min at 14000 rpm, 4°C.

5. Aliquot the supernatant and dilute the samples with 4% milk in WBT to a final concentration of 1% milk. For inhibition studies, different concentrations of competitor should be added to the different aliquots and the mixture should be equilibrated for 1 h at room temperature prior to step 6.
6. Split the samples to at least duplicates and apply them to the blocked plate. Let the protein bind to the target molecule by slowly shaking for maximally 30 min.
7. Wash the plate rigorously 5-10 times with WBT.
8. Elute bound protein with 100 µl of a 0.1 M solution of triethylamine (10 min at room temperature).
9. Transfer the eluates into scintillation tubes containing 5 ml scintillation solution "OptiPhase2".
10. Quantify the radioactivity in a scintillation counter.

IV. Comments

In this section we would like to state some observations we made during the years of performing ribosome display and give a summary how ribosome display can be used for directed evolution experiments.

Selection from Naïve Libraries

Naïve libraries, in our hands synthetic antibody libraries or designed repeat protein libraries, are a difficult challenge for selection experiments. The task is to select the specific binders, which are few in numbers, out of a very large number of non-binders. The outcome of such experiments is not only dependent on the presence of high affinity binders in the library, but also on the behavior, or "stickiness", of the rest of the library population. In other words, selection must be directed towards specific binding, in contrast to non-specific binding. This can be achieved with experimental tricks, such as introducing a pre-panning step, or by trying to reduce the stickiness of the library population. For the selection with naïve scFv libraries it has turned out that six selection rounds were necessary to obtain specific binders (Hanes et al., 2000b). The designed ankyrin repeat protein libraries routinely yield binders after only three to four rounds. We suspect this might be due to the fact that these repeat proteins, which are extremely well expressed in *E. coli*, fold well and are very stable, are also displayed better *in vitro* than scFv fragments, which are intrinsically somewhat more aggregation prone.

Affinity Maturation of Binders

In vitro evolution, the alternation of diversification and selection, is a powerful strategy to improve proteins. Ribosome display is an ideal platform to perform such experiments. It allows very fast selection cycles and the PCR step is ideal to generate diversity in between the selection rounds. This diversification of the selected pools by random mutations increases the sampled sequence space. Error prone PCR, e.g. using high Mn^{2+} concentrations (Leung et al., 1989), imbalanced dNTP concentrations (Cadwell and Joyce, 1994) or nucleotide analogues (Zaccolo and Gherardi, 1999; Zaccolo et al., 1996), is one strategy often applied to achieve diversification. Another powerful strategy to create diversity is DNA shuffling on the selected pools in between the rounds (Stemmer, 1994).

When compared to a selection experiment from a naïve library the challenge in affinity maturation is different. The applied library will usually be created from a single clone or a pool of clones, which are already good binders. Since we do not select for binding as such but for better binding, we need an adjustable selection pressure (Jermutus et al., 2001). In principle, there are two strategies to select tight binders out of a pool of binders.

The first affinity maturation strategy is to supply very little of the immobilized antigen. The concept is that the binders all compete with each other and at equilibrium, the tight binders keep the binding spaces occupied. In practice, however, there are several complications to this concept. At extremely low target molecule concentrations, the relative proportion of unspecific binding sites gets very high (BSA, neutravidin), so that great care has to be taken to avoid unspecific binders. Also, if medium affinity binders outnumber the tight binders, the enrichment will be very slow. Finally, if binding is very tight, the equilibrium is reached exceedingly slow.

The second very successful strategy for affinity maturation is off-rate selection. The assumption fundamental to off-rate selection is that the on-rate of most protein-protein interactions is in the range of 10^5-10^6 $M^{-1}sec^{-1}$ (Wodak and Janin, 2002) and protein ligand interactions in the range of 10^6-10^7 $M^{-1}sec^{-1}$, such that the affinity is largely governed by the off-rate. By first incubating the ribosome-displayed polypeptide with biotinylated target molecule (typically for one hour) followed by the addition of a large excess of non-biotinylated target (1000-fold excess), the selection pressure is governed by the dissociation of the binders from the biotinylated target. While tight binders will remain bound to the biotinylated target, others will dissociate and then rebind to the excess of unbiotinylated target. The incubation time equals the selection pressure and should be adjusted to the expected off-rate. If the recovery is poor, it is often useful to include a non-selective round to enrich the binders.

Evolution of Properties other than High Affinity Binding

The selection strategies for molecular properties other than high affinity binding have been summarised in the Introduction. A library can be evolved for these properties, just as it can be for affinity. It exceeds the scope of this chapter to discuss each strategy in detail, and there are many more possibilities, which have not been experimentally explored.

Acknowledgment

The authors would like to thank all former and present members of the Plückthun laboratory involved in ribosome display who helped in developing the present protocol.

V. References

Amstutz, P., Forrer, P., Zahnd, C. and Plückthun, A. (2001). *In vitro* display technologies: novel developments and applications. *Curr. Opin. Biotechnol.* 12, 400-405.

Amstutz, P., Pelletier, J. N., Guggisberg, A., Jermutus, L., Cesaro-Tadic, S., Zahnd, C. and Plückthun, A. (2002). *In vitro* selection for catalytic activity with ribosome display. *J. Am. Chem. Soc.* 124, 9396-9403.

Cadwell, R. C. and Joyce, G. F. (1994). Mutagenic PCR. *PCR Methods Appl.* 3, 136-140.

Chen, H. Z. and Zubay, G. (1983). Prokaryotic coupled transcription-translation. *Methods Enzymol.* 101, 674-690.

Hanes, J., Jermutus, L. and Plückthun, A. (2000a). Selecting and evolving functional proteins *in vitro* by ribosome display. *Methods Enzymol.* 328, 404-430.

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. and Pluckthun, A. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries. *Proc. Natl Acad. Sci. USA* 95, 14130-14135.

Hanes, J. and Plückthun, A. (1997). *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA* 94, 4937-4942.

Hanes, J., Schaffitzel, C., Knappik, A. and Plückthun, A. (2000b). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* 18, 1287-1292.

He, M. and Taussig, M. J. (1997). Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for *in vitro* display and evolution of antibody combining sites. *Nucleic Acids Res.* 25, 5132-5134.

Irving, R. A., Coia, G., Roberts, A., Nuttall, S. D. and Hudson, P. J. (2001). Ribosome display and affinity maturation: from antibodies to single V- domains and steps towards cancer therapeutics. *J. Immunol. Methods* 248, 31-45.

Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. and Plückthun, A. (2001). Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl Acad. Sci. USA* 98, 75-80.

Kushner, S. R. (2002). mRNA decay in *Escherichia coli* comes of age. *J. Bacteriol*. 184, 4658-4665.

Laminet, A. A. and Plückthun, A. (1989). The precursor of beta-lactamase: purification, properties and folding kinetics. *Embo J.* 8, 1469-1477.

Lesley, S. A. (1995). Preparation and use of *E. coli* S-30 extracts. *Methods Mol. Biol.* 37, 265-278.

Leung, D. W., Chen, E. and Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1, 11-15.

Matsuura, T. and Plückthun, A. (2003). Selection based on the folding properties of proteins with ribosome display. *FEBS Lett.* 539, 24-28.

Mattheakis, L. C., Bhatt, R. R. and Dower, W. J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl Acad. Sci. USA* 91, 9022-9026.

Pratt, J. M. (1984). Coupled transcription-translation in prokaryotic cell-free systems. In *Current protocols* (Hemes, B. D. and Higgins, S. J., eds.), pp. 179-209. IRL Press, Oxford.

Schatz, P. J. (1993). Use of peptide libraries to map the substrate specificity of a peptide- modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Biotechnology (New York)* 11, 1138-1143.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315-1317.

Stemmer, W. P. (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389-391.

Takahashi, F., Ebihara, T., Mie, M., Yanagida, Y., Endo, Y., Kobatake, E. and Aizawa, M. (2002). Ribosome display for selection of active dihydrofolate reductase mutants using immobilized methotrexate on agarose beads. *FEBS Lett.* 514, 106-110.

Wade, H. E. and Robinson, H. K. (1966). Magnesium ion-independent ribonucleic acid depolymerases in bacteria. *Biochem. J.* 101, 467-479.

Wodak, S. J. and Janin, J. (2002). Structural basis of macromolecular recognition. *Adv. Prot. Chem.* 61, 9-73.

Zaccolo, M. and Gherardi, E. (1999). The effect of high-frequency random mutagenesis on *in vitro* protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* 285, 775-783.

Zaccolo, M., Williams, D. M., Brown, D. M. and Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* 255, 589-603.

Zubay, G. (1973). *In vitro* synthesis of protein in microbial systems. Annu. Rev. Genet. 7, 267-287.

Addresses of Suppliers:

Avestin Europe GmbH
Weinheimer Str. 64b
D68309 Mannheim
Germany

Corning B.V.
Life Sciences
Koolhovenlaan 12
NL-1119 NE Schiphol-Rijk
The Netherlands

EMD Biosciences, Inc.
CALBIOCHEM®
10394 Pacific Center Court
San Diego, California 92121
USA

Eppendorf AG
Barkhausenweg 1
D-22339 Hamburg
Germany

FERMENTAS UAB
V.Graiciuno 8,
Vilnius 2028
Lithuania

Fluka
Industriestrasse 25
CH-9471 Buchs
Switzerland

Hoffmann-La Roche Ltd, Diagnostics Division
Grenzacherstrasse 124
CH-4070 Basel
Switzerland

Invitrogen AG
Elisabethenstrasse 3
Postfach 533
CH-4019 Basel
Switzerland

Merck KGaA
Frankfurter Str. 250
D-64293 Darmstadt
Germany

Millipore
290 Concord Rd.
Billerica, MA 01821
USA

Molecular BioProducts, Inc.
9880 Mesa Rim Road
San Diego, CA 92121-2979
USA

New England Biolabs, Inc.
32 Tozer Road
Beverly, MA 01915-5599
USA

NUNC
Kamstrupvej 90
Postbox 280
DK-4000 Roskilde
Denmark

PerkinElmer
45 William Street
Wellesley, MA 02481-4078
USA

Pierce Biotechnology
P.O. Box 117
Rockford, IL 61105
USA

QIAGEN AG
Auf dem Wolf 39
CH-4052 Basel
Switzerland

Spectrum Laboratories, Inc.
America and Asia Pacific
18617 Broadwick Street
Rancho Dominguez, CA 90220
USA

Stratagene Corporate Office
11011 N. Torrey Pines Road
La Jolla, CA 92037
USA

Chapter 4

High-affinity Binders Selected from Designed Ankyrin Repeat Protein Libraries

Contents

ARTICLES

nature biotechnology

High-affinity binders selected from designed ankyrin repeat protein libraries

H Kaspar Binz[1,2], Patrick Amstutz[1,2], Andreas Kohl[1,2], Michael T Stumpp[1], Christophe Briand[1], Patrik Forrer[1], Markus G Grütter[1] & Andreas Plückthun[1]

We report here the evolution of ankyrin repeat (AR) proteins *in vitro* for specific, high-affinity target binding. Using a consensus design strategy, we generated combinatorial libraries of AR proteins of varying repeat numbers with diversified binding surfaces. Libraries of two and three repeats, flanked by 'capping repeats,' were used in ribosome-display selections against maltose binding protein (MBP) and two eukaryotic kinases. We rapidly enriched target-specific binders with affinities in the low nanomolar range and determined the crystal structure of one of the selected AR proteins in complex with MBP at 2.3 Å resolution. The interaction relies on the randomized positions of the designed AR protein and is comparable to natural, heterodimeric protein-protein interactions. Thus, our AR protein libraries are valuable sources for binding molecules and, because of the very favorable biophysical properties of the designed AR proteins, an attractive alternative to antibody libraries.

Repeat proteins are ubiquitous binding molecules fundamental to many biological processes[1–3]. Their modular architecture is presumably the key to their evolutionary success[4]. Repeat proteins are characterized by consecutive homologous structural units (repeats), which stack to form an elongated protein domain with a continuous hydrophobic core[3]. In principle, this architecture allows their binding specificities to evolve not only by point mutations but also by insertion, deletion or shuffling of repeats[5]. This evolutionary strategy might enable repeat proteins to acquire new functions by adjusting their surface without jeopardizing their overall topology. AR proteins are one prominent repeat protein family illustrating the binding versatility of repeat proteins. They occur throughout all phyla and mediate protein-protein interactions in the nucleus or cytoplasm, or while anchored to the membrane or when secreted into the extracellular space[6]. AR proteins are built from stacked, 33 amino acid repeats, each forming a β-turn that is followed by two antiparallel α-helices and a loop reaching the β-turn of the next repeat[7]. In most known complexes, the β-turn and the first α-helix mediate the interactions with the target, and different numbers of adjacent repeats are involved in binding[7]. The reported target binding affinities of natural AR proteins are in the low nanomolar range[8,9].

In biotechnology and biomedical research, antibodies and fragments thereof are the most widely used specific, high-affinity binding molecules. Antibodies can be generated against essentially any target either by immunization or by using natural or rationally designed antibody libraries *in vitro*[10,11]. Yet, many antibodies have relatively low expression yields, a tendency to aggregate and a dependence on disulfide bonds for stability. An ideal alternative protein scaffold would have none of these drawbacks while still exhibiting the same affinity and specificity as antibodies. Previous attempts to generate alternative binding molecules relied on either loop or surface randomization of protein scaffolds, which are typically small and always fixed in dimension[12–14].

We present here the results of a different strategy[15] to design and select alternative protein scaffolds, relying on the modularity of AR proteins (**Fig. 1**). We generated combinatorial libraries of consensus-designed AR proteins of varying sizes (that is, varying repeat numbers) with randomized potential interaction surfaces. Unselected library members are very well expressed, soluble, thermodynamically stable and show the typical AR domain fold[16,17]. Here, we show the successful selection of binding molecules from these libraries, proving that our design strategy works. We selected specific binders with high affinities for the *Escherichia coli* maltose binding protein (MBP) and two eukaryotic mitogen-activated protein kinases (MAPKs). The crystal structure of one of the selected binders in complex with MBP was determined, revealing atomic level insights into the mode of target interaction of this class of designed binding molecules.

RESULTS

Designed AR protein libraries

We designed a consensus AR module consisting of six diversified potential interaction residues (which can be any amino acid except cysteine, glycine and proline) and 27 framework residues (26 are fixed and one is allowed to be asparagine, histidine or tyrosine). This module was designed from sequence alignments and structural analyses (**Fig. 1**)[16]. The randomized potential interaction residues are located in the β-turn and the first α-helix of the AR module. We cloned varying numbers of this repeat module between capping repeats, which are special terminal repeats of AR domains shielding the hydrophobic

[1]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [2]These authors contributed equally to this work. Correspondence should be addressed to A.P. (plueckthun@bioc.unizh.ch) or M.G.G. (gruetter@bioc.unizh.ch).

Published online 18 April 2004; doi:10.1038/nbt962

ARTICLES

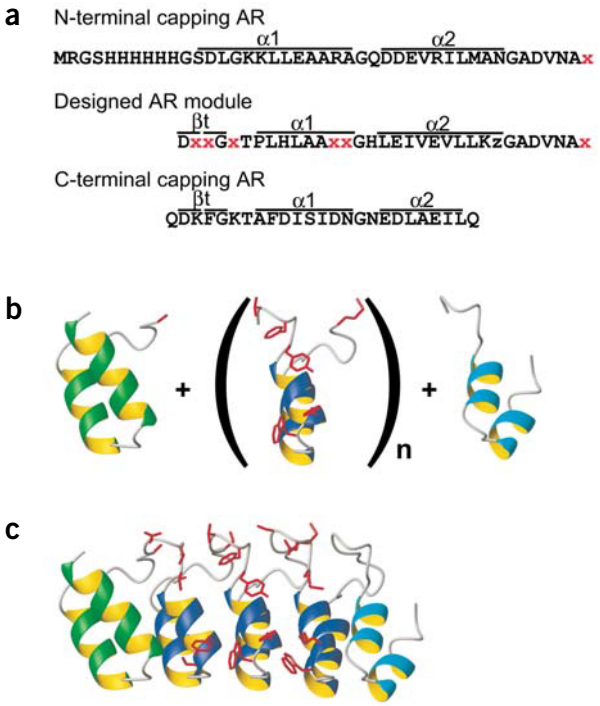


Figure 1 Construction of designed AR protein libraries. (**a**) Sequences of the N-terminal capping AR, the designed AR module and the C-terminal capping AR. The secondary structure elements are indicated above the sequences. The designed AR module consists of 26 defined framework residues, six randomized potential interaction residues (red x, any of the 20 natural amino acids except cysteine, glycine or proline) and one randomized framework residue (z, any of the amino acids asparagine, histidine or tyrosine). The designed AR module was derived via sequence and structure consensus analyses[16]. (**b**) Schematic representation of the library generation of designed AR proteins. Note that this assembly is represented on the protein level, whereas the real library assembly is on the DNA level. By assembling an N-terminal capping AR (green), varying numbers of the designed AR module (blue) and a C-terminal capping AR (cyan), combinatorial libraries of designed AR proteins of different repeat numbers were generated (side chains of the randomized potential interaction residues are shown in stick-mode in red). (**c**) Ribbon representation of the selected MBP binding AR protein off7 (colors as in **b**). This binder is derived from a library consisting of a N-terminal capping AR, three designed AR modules and a C-terminal capping AR. This figure was made with MolMol[49].

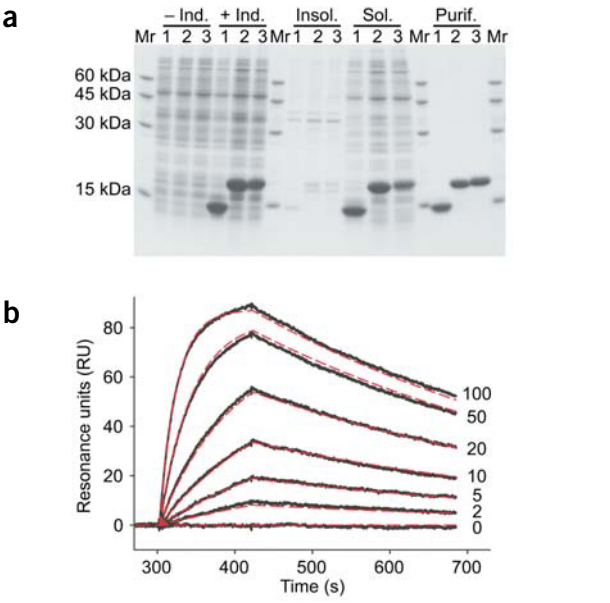


Figure 2 Expression, purification and SPR analysis of selected AR proteins. (**a**) Expression and purification of the selected MBP binders mbp3_16 (1), off7 (2) and mbp3_5 (3). At OD_{600} = 0.6, the noninduced (– Ind.) cultures were induced with 0.5 mM IPTG and grown for 4 h at 37 °C (+ Ind.). After cell lysis, the AR proteins are in the soluble fraction (Sol.). Insol., insoluble fraction. The proteins were then purified in a single IMAC purification step (Purif.). (**b**) BIAcore analysis of off7. Different concentrations of off7 (0, 2, 5, 10, 20, 50 and 100 nM) were applied to a flow cell with immobilized MBP for 2 min, followed by washing with buffer flow. The global fit is indicated in the figure by red dashed lines (see **Table 1** for the extracted kinetic data).

core (**Fig. 1**). We thereby increased the size of the potential interaction surface and potentiated its diversity. This strategy provided combinatorial libraries of designed AR proteins with distinct repeat numbers. The libraries were named N2C and N3C, indicating proteins had an N-terminal capping repeat, two and three, respectively, designed (and randomized) AR modules and a C-terminal capping repeat (**Fig. 1**). We used the N2C and N3C libraries for the selections, because AR proteins of this length are very abundant in nature[6]. Unselected members of these libraries were expressed in soluble form at about 200 mg/l in *E. coli* shake flask cultures. These proteins were monomeric, showed circular dichroism (CD) spectra indistinguishable from natural AR proteins[16] and the AR fold was confirmed by a crystal structure of an unselected library member[17]. Similar to designed proteins in other consensus repeat protein studies[18,19], our designed AR proteins showed high thermodynamic stability during unfolding induced by heat[16] and denaturants[17]. Hence, the consensus-designed AR proteins are stable scaffolds with large and modular potential interaction surfaces. The theoretical diversities of these libraries are $5.2{\cdot}10^{15}$ (N2C) and $3.8{\cdot}10^{23}$ (N3C)[16]. The DNA libraries used in the selections contained at least 10^{10} individual members each as estimated from the amount of ligated library DNA. The library diversities were further increased in subsequent PCR cycles.

Ribosome-display selection against MBP

We chose *E. coli* MBP as the first target protein for evaluating our libraries, because it can be obtained in large amounts in pure form and because its structure is known (Protein Data Bank (PDB) entry 1LLS)[20]. We did the ribosome-display selections[21] with biotinylated MBP bound to neutravidin in microtiter plates. An enrichment of binders was observed after the second selection round both for the N2C and the N3C libraries. We performed a total of four to five selection rounds before analyzing single, selected library members.

We screened individual selected AR proteins for MBP binding by an enzyme-linked immunosorbent assay (ELISA) using crude *E. coli* extracts. Of 60 N3C AR proteins screened, 18 gave a specific signal (signal/background ≥ 10), compared to 4 of 56 N2C molecules. Sequencing of the 18 MBP binding N3C AR proteins revealed that they could be divided into at least three sequence groups (see **Supplementary Fig. 1** online). However, identical clones were never found and considerable diversity was left, indicating that an even more stringent selection pressure could be applied. For the N2C clones the sequence analysis was less conclusive because of the limited data set. Nevertheless, some repeats of N2C molecules showed striking sequence similarities to repeats of the selected N3C proteins. In both

Table 1 Kinetic binding data of selected clones determined by surface plasmon resonance

Target	Clone name (length)	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_D [M]
MBP	off7 (N3C)	$4.2 \cdot 10^5$	$1.9 \cdot 10^{-3}$	$4.4 \cdot 10^{-9}$
	mbp3_5 (N3C)	$2.0 \cdot 10^5$	$4.4 \cdot 10^{-3}$	$22 \cdot 10^{-9}$
	mbp3_16 (N2C)	$6.0 \cdot 10^5$	$1.0 \cdot 10^{-2}$	$17 \cdot 10^{-9}$
JNK2	JNK2_2_3 (N2C)	$9.7 \cdot 10^5$	$2.0 \cdot 10^{-3}$	$2.1 \cdot 10^{-9}$
p38	p38_2_3 (N2C)	$9.5 \cdot 10^5$	$3.5 \cdot 10^{-3}$	$3.7 \cdot 10^{-9}$

N2C and N3C molecules, aromatic residues appeared frequently and thus seem to be important in the selected sequences (see **Supplementary Fig. 1** online).

Framework mutations were randomly scattered and occurred at low frequencies but were present in all molecules (on average, 3.8 amino acid mutations per N3C and 2.4 mutations per N2C molecule; see **Supplementary Fig. 1** online). Each library member went through at least 450 PCR cycles (library generation and selection), which might explain this finding. The mutations are mostly of a conservative nature, and thus the fold of the AR domains is most probably not affected by the alteration of the framework (see description of the crystal structure below).

Selected AR proteins show high affinity and specificity

The selected AR proteins were expressed at high levels in soluble form in the cytoplasm of *E. coli* (up to 200 mg/l) and purified to homogeneity by a single immobilized metal ion affinity chromatography (IMAC) purification step (**Fig. 2a**). We screened 21 clones (4 N2C, 17 N3C) by surface plasmon resonance (SPR). Using the purified AR proteins at 1 μM, we first compared the on- and off-rates of MBP binding. Three AR proteins with slow off-rates were analyzed at multiple concentrations and evaluated with a global kinetic fit (N2C: mbp3_16; N3C: off7 and mbp3_5) (**Fig. 2b** and **Table 1**). off7, a selected N3C molecule, had the highest affinity for MBP (K_D= 4.4 nM). The N2C molecule mbp3_16 had a dissociation constant of K_D= 17 nM. Hence, both N2C and N3C molecules can be selected to bind MBP with high affinity. Clones that went through five selection rounds had higher affinities for MBP than clones that were selected through four rounds. In SPR experiments, the N3C library member off7 was specific and did not cross-react with phage lambda protein D[22], streptavidin or the aminoglycoside-3′-phosphotransferase APH(3′)-IIIa, a bacterial kanamycin resistance protein[23].

To further investigate specificity, ELISA experiments were done with purified MBP-binding AR proteins (**Fig. 3**). mbp3_16, off7 and mbp3_5 are specific for MBP and do not interact with phage lambda protein D[22], APH[23] or neutravidin (**Fig. 3a**). In a competition ELISA experiment, the binding of off7 to immobilized MBP could be inhibited by preincubation with free MBP (**Fig. 3b**). The affinity estimated from these experiments was consistent with the SPR measurements (50% inhibition at 10 nM). The unselected N3C AR library member E3_5[16,17] did not interact with MBP, indicating that the designed AR domain scaffolds per se do not bind MBP (**Fig. 3b**).

At a concentration of 15 μM, off7 is monomeric (as indicated by size exclusion chromatography) and shows a CD spectrum identical to that of E3_5 (ref. 16), which has an AR domain fold[17] (data not shown).

Selection of specific high-affinity MAPK binders

To further evaluate the potential of our AR protein libraries, we chose the eukaryotic protein kinases JNK2 and p38 as our next target

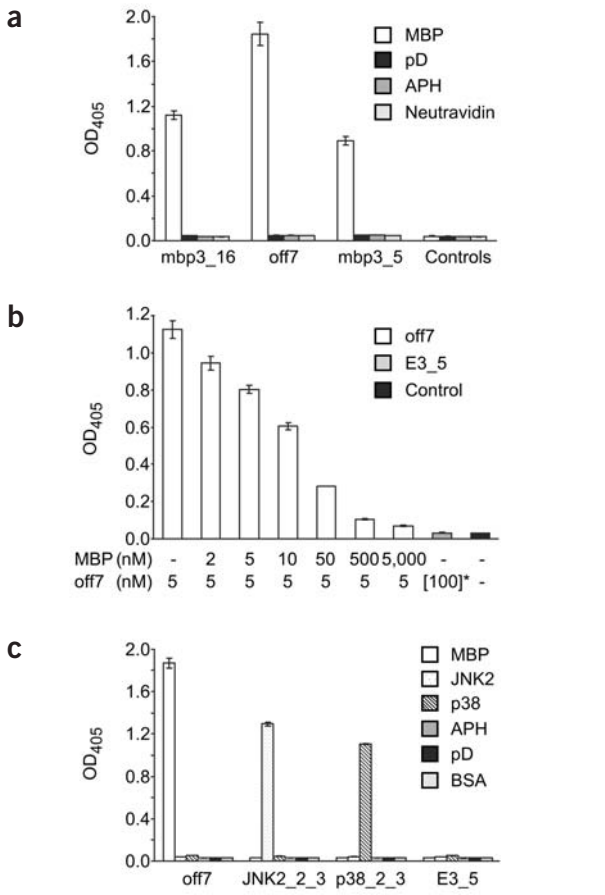


Figure 3 ELISAs with selected AR proteins. (**a**) Specificity of MBP binders. The interaction of the proteins mbp3_16, off7 and mbp3_5 (each 50 nM; control with no AR protein) with immobilized MBP, pD, APH and neutravidin is shown. (**b**) Competition ELISA illustrating the interaction between the selected AR protein off7 and MBP. off7 (5 nM) was incubated with varying concentrations of free MBP before binding on immobilized MBP. The binding to MBP of off7 can be specifically inhibited by increasing concentrations of free MBP in solution. An unselected AR protein of the N3C library (*) showed no interaction with MBP (100 nM of E3_5)[16,17], giving a signal identical to that of the control (no AR protein on immobilized MBP). (**c**) Specificity comparison of an MBP, a JNK2 and a p38 binder. The interaction of 100 nM each of the proteins off7 (binds MBP), JNK2_2_3 (binds JNK2), p38_2_3 (binds p38) and E3_5 (unselected N3C library member) with MBP, JNK2, p38, APH, pD and BSA is shown. Note that in all representations the background binding of the detection antibodies has not been subtracted.

proteins (see ref. 24 and references therein). We did a total of four ribosome-display selection rounds with the N2C library before comparing single, selected library members. Screening 15 clones each, we obtained ten ELISA-positive JNK2 binders and ten ELISA-positive p38 binders. The sequence of one representative member for each of these target-specific groups is given in **Supplementary Figure 1** online. These MAPK binders share many features of the selected MBP binders. They have affinities in the low nM range (**Table 1**); they can be expressed at high levels in soluble form in the cytoplasm of *E. coli* and purified to homogeneity by a single IMAC purification step (data not shown); their randomized positions are enriched in aromatic amino acids (see **Supplementary Fig. 1** online). To investigate their target specificity, we did ELISA experiments with purified, selected AR

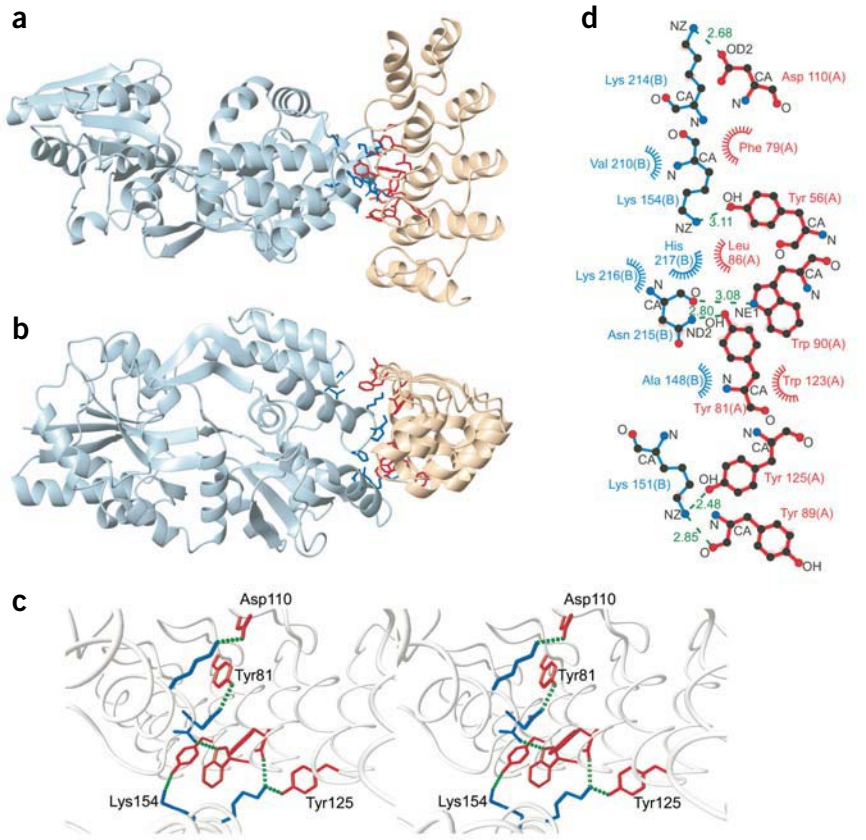


Figure 4 Crystal structure of the designed AR protein off7 in complex with MBP. (**a**,**b**) Two perpendicular views of the complex are shown. MBP is on the left (blue), off7 on the right (ochre). The interaction residues are highlighted in stick-mode in red (off7) and blue (MBP), respectively. (**c**) A close stereo view on the H-bond pattern in similar view as in **b**. Note that in this representation only residues involved in H-bonds (green dashed lines) are shown. For orientation, some residues involved in H-bonding are labeled. **Figure 4a–c** were made with MolMol[49]. (**d**) Ligplot[48] representation of the interaction between MBP (chain B, blue) and off7 (chain A, red). H-bonds (in green) including the H-bond distances as well as residues and atoms involved in hydrophobic contacts (indicated by red or blue rays) are shown.

proteins (**Fig. 3c**). All binders were highly specific for the target proteins that were used to select them. Most importantly, the AR proteins allow perfect discrimination between the homologous MAPKs (51% identity and 59% similarity between JNK2 and p38 on the amino acid level in our format).

Structure determination of off7 in complex with MBP

To validate our AR randomization scheme and to analyze the selected interaction at the atomic level, we determined the crystal structure of one binder (off7) in complex with MBP (see Methods). The phasing problem was solved by molecular replacement using E3_5, a designed N3C AR protein[17] as a search model without using the phases of the larger MBP. Hence, when used in cocrystallization studies, AR proteins may serve as valuable tools to obtain first phases and finally the structure of its binding partner. The results of the data collection and refinement are shown in **Supplementary Table 1** online. As can be seen in the crystal structure (**Figs. 4** and **5**), the AR protein binds the open form of MBP[20] creating an elongated complex. For the AR protein, clear electron density starts at Ser12, but no or only very weak electron density was observed for the N-terminal His_6 tag. The electron density is clear until the second-to-last amino acid (Leu168). In the complex, the AR protein has the typical AR domain fold and is highly similar to the known structure of the designed AR protein E3_5 (root-mean-square deviation of the C_α atoms (r.m.s.d.$_{C\alpha}$) < 1 Å)[17]. The main differences between the two structures are found in the β-turn region of repeat module 1 (second repeat), where the C_α chain shows a maximal r.m.s.d.$_{C\alpha}$ of 2.1 Å. The H-bond network, which most probably stabilizes the β-turn region[17], is conserved and undisturbed. off7 has three framework mutations (K16R, N74D and H125Y) not present in the library design, but they are conservative and in accordance with the AR domain fold, and only Y125 is involved in MBP binding (see below).

MBP was found in the open conformation with no ligand bound[20]. A superposition on PDB entry 1LLS[20] shows very few differences (r.m.s.d.$_{C\alpha}$ < 0.9 Å). The N-terminal His_6 tag and the C terminus of MBP are not defined in the electron density. Clear density extends from Gly19 to Gly387.

Analysis of the interaction of off7 with MBP

The interaction of the AR protein off7 with MBP was analyzed as described in the Methods section. The only direct interaction between off7 and MBP in the crystal lattice is the selected heterodimer interface (**Fig. 4**). Crystal packing contacts are mainly between adjacent off7 molecules and between adjacent MBP molecules. The heterodimer interface is formed by the concave randomized surface of the AR protein off7 (611 Å^2 buried surface) and a slightly larger convex surface on the MBP (656 Å^2 buried surface), resulting in a total buried surface area of 1,267 Å^2 (**Fig. 5**). The off7/MBP complex is further characterized by six H-bonds and a planarity index of 2.1[25]. The details of the interaction are listed in **Table 2** and in **Supplementary Table 2** online.

The interaction of off7 involves residues from all three randomized repeat modules, although the randomized repeat modules 2 and 3 (constituting repeats 3 and 4 in the protein because of the capping repeats) contribute more to the binding than the randomized repeat module 1 does. In total, 9 out of 18 randomized potential interaction residues are involved in the binding to MBP. The interface is characterized by a large number of aromatic residues, which account for 73% of the buried surface area on off7 (**Fig. 4** and **Supplementary Table 2** online). Among these residues, the tyrosines cover 28% of the buried surface area and are involved in four H-bonds (Tyr56, Tyr81, Tyr89 and Tyr125). In the interface (**Fig. 4**), tyrosines have a dual role being both H-bond formers and hydrophobic contact mediators. Besides the four tyrosines, Trp90 and Asp110 also form H-bonds. Three framework residues (Leu86, Asp110 and Tyr125) form part of the interaction surface. Two (Asp110 and Tyr125) form H-bonds to MBP (**Fig. 4**) and the third (Leu86) is engaged in hydrophobic interactions. Interestingly, Tyr125 is a framework mutation (H125Y). With the

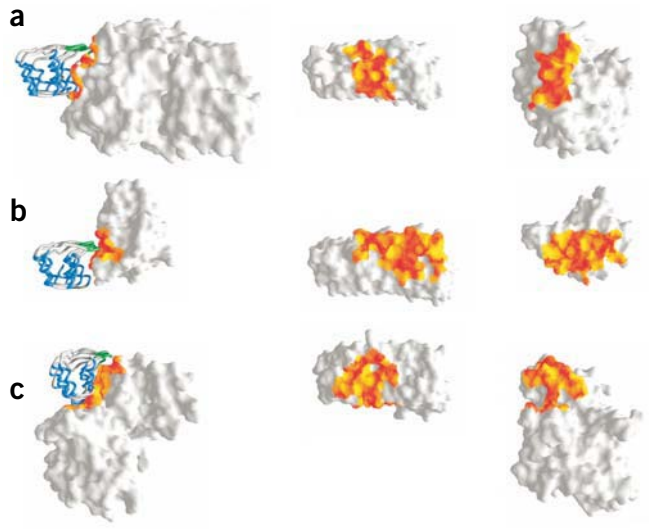


Figure 5 Open sandwich illustrations of the interaction surfaces of AR proteins and their targets. (**a**–**c**) GRASP[50] shape complementarity representations of the interactions between off7 and MBP (shape complementarity, 0.739) (**a**), GABPβ1 and GABPα (PDB entry 1AWC; shape complementarity, 0.665) (**b**), and $p18^{INK4c}$ and CDK6 (PDB entry 1G3N; shape complementarity, 0.688) (**c**), respectively. The complex is shown on the left with the AR proteins in a backbone worm representation (α-helices in blue, β-turns in green) and the target protein in a surface representation. The open sandwich surface representations are shown in the middle (AR proteins) and on the right (targets). The contact areas are stained according to the shape complementarities from orange (low) to red (high).

exception of the high percentage of aromatic amino acids, the residue composition of the off7 interface is comparable to other AR protein complexes. However, the small and functionally biased set of only seven structurally analyzed natural AR protein complexes limits the comparison to their average amino acid composition in the binding interface.

The interaction surface of MBP is located on helices H7, H9 and H17 (**Fig. 4** and **Supplementary Table 2** online). Four lysines (Lys151, Lys154, Lys214 and Lys216) are involved in four H-bonds and together form approximately 60% of the buried surface area on the MBP surface (**Fig. 4**). Every lysine of MBP is in contact with a tyrosine of off7 resulting in three H-bonds and five hydrophobic contacts (see **Supplementary Table 2** online). In general, the surface of the MBP is rather negatively charged except for the spot where off7 binds, where there is a positively charged surface patch formed by the four lysines.

A comparison to natural protein-protein interactions

We compared the off7/MBP complex to all available complex crystal structures of AR proteins, to natural heterodimer complexes, to antibody-antigen complexes and to an affibody complex (see Methods section and **Table 2**). The off7/MBP interaction is comparable to natural protein-protein interactions[25]. The 611 $Å^2$ buried surface area, one H-bond per 100 $Å^2$ buried surface area and the planarity value of 2.1 of off7, are all within the normal parameters of heterodimer complexes (**Table 2**)[25]. The buried surface area is at the lower limit of antibody-antigen complexes[26], whereas the affinity is comparable to high affinity monovalent antibody-antigen binding. Otherwise, all values seem to fit the standard parameters for antibody-antigen complexes quite well, but are also similar to those of heterodimer complexes. Natural AR protein interactions can be very diverse in size, composition and in the structural elements involved (**Fig. 5** and **Table 2**). Our randomization scheme of the β-turn and the first α-helix of the designed AR modules was based on crystal structures of natural AR complexes, where it was clear that the AR scaffold is directly used for binding[16]. Because 50% of the randomized target interaction residues that were selected do indeed interact with MBP, the interaction mode is very similar to that of natural AR proteins, such as the GABPβ1 (**Fig. 5**), which partly inspired our library design[16]. In comparison to the natural AR proteins, off7 has a slightly smaller buried surface area but a higher H-bond density (**Table 2**).

Recently the structure of an affibody, another designed binding molecule, based on the staphylococcal three-helix bundle protein A, in complex with its target protein was published[27]. The complex shows a slightly larger buried surface area than the off7/MBP complex but with a comparable number of H-bonds (**Table 2**). The affibody has a thousand-fold lower affinity for its target[28] than the AR protein off7 does for MBP ($K_D = 6\ \mu M$ vs. $K_D = 4.4$ nM, respectively). This is probably because this affibody is in a molten globule state and assumes a defined structure only upon binding[29], leading to a loss of entropy that reduces the overall free energy of binding and thus the observed affinity.

DISCUSSION

We designed AR protein libraries of varying repeat numbers using a consensus design strategy[15–17]. Here we show the successful selection of binding molecules from these libraries. The properties of the designed AR proteins perfectly match the criteria for alternative scaffolds. They are expressed at a high level in soluble form, are monomeric and do not contain any cysteines[16]. Unlike previously presented scaffolds (for reviews see refs. 12,13), which typically use either randomized loops or a randomized surface on a given protein scaffold for binding, AR proteins use both β-turns and a randomized surface. Most importantly, they are not restricted in dimension because of their modular architecture (**Fig. 1**). Thus, the interaction surface can be adapted by adding more repeat modules. The favorable properties of the molecules in the starting libraries seem to positively influence both the speed of selection and its outcome. In only four to five ribosome-display selection rounds, we were able to enrich the pool in specific, high-affinity protein binders from the N2C and N3C libraries, which appears to be faster than ribosome-display selections from antibody single-chain Fv libraries[30].

The selected AR proteins retain the advantageous properties of the designed AR proteins, being expressed in high amounts in soluble form and free of cysteines (**Fig. 2**). The selected clones specifically recognize the target protein against which they were selected and do not cross-react with other proteins as shown by ELISA (**Fig. 3**). The affinities of the selected clones are in the low nanomolar range (**Table 1** and **Fig. 2**), the association rates are in the typical range for protein-protein interactions (that is, 10^5–$10^6\ M^{-1}s^{-1}$)[31] and the dissociation rates are in the range of 10^{-2} to $2{\cdot}10^{-3}\ s^{-1}$ (**Table 1**). We anticipate that these off-rates can probably be improved further by a diversification step followed by an off-rate ribosome-display selection round[32]. Such a diversification step could involve not only classical error-prone PCR, but also, more interestingly, strategies purely amenable to repeat proteins such as repeat shuffling or repeat elongation.

The selected AR protein sequences contain a high number of aromatic residues (see **Supplementary Fig. 1** online) and in the case of off7, seven aromatic amino acids are involved in MBP binding (**Fig. 4**), including four prominent tyrosines (**Supplementary Table 2** online). Interestingly, high tyrosine content has also been noted in antibody binding sites[26]. The dual interaction role of tyrosine as H-bond former and hydrophobic contact mediator (see Results)[33] is probably the

Table 2 Comparison between AR protein complexes and other protein-protein interactions

PDB entry	Resolution	ΔASA[a] in Å^2	No. H-bonds	No. H-bonds/100 Å^2 ΔASA[a]	No. of salt bridges	Planarity	No. of bridging H_2O
1awc	2.2	853.62	5	0.58	0	2.30	3
1bi7	3.4	1205.5	7	0.29	4	2.50	0
1blx	1.9	845.4	11	1.30	1	2.50	12
1g3n	2.9	843.3	12	1.42	1	2.20	0
1ikn	2.3	763.96	2	0.26	n.d.	2.73	0
1oy3	2.1	1593.9	7	0.44	n.d.	4.01	10
1ycs[b]	2.2	713.7	7	1.00	1	2.60	1
1svx (off7/MBP)	2.3	611.2	6	1.00	0	2.10	0
1lp1 (Affibody)	2.3	848.9	6	0.71	0	2.00	2
Antibody-antigen complexes[c]	–	777 ± 135	–	1.1 ± 0.5	–	2.2 ± 0.4	–
Heterodimeric protein-protein complexes [c]	–	983 ± 582	–	1.1 ± 0.5	–	2.8 ± 0.9	–

[a]Surface area per molecule occluded upon complex formation. [b]1ycs uses a different binding surface than the other AR proteins in this table. [c]According to ref. 25.

reason for this accumulation. In addition, aromatic residues are generally enriched in protein-protein interaction interfaces[25].

Apart from the high content of aromatic residues, the crystal structure of the off7/MBP complex reveals an interaction interface that is comparable to that found in natural heterodimer and antibody-antigen complexes (**Table 2**)[25]. It also shows that the AR protein binds its target with the randomized amino acids (**Fig. 4** and **Supplementary Table 2** online), hence validating our randomization scheme. In the crystal structure of the complex, off7 shows the typical AR domain fold with a high similarity to the unselected N3C library member E3_5 (ref. 17). In solution, the uncomplexed, monomeric off7 shows a CD spectrum virtually identical to that of E3_5 (ref. 16). Thus, the backbone of off7 does not seem to be rearranged in a substantial way upon binding to MBP but rather seems to interact in a key-to-lock mechanism. Such a rigid-body interaction may be advantageous both for affinity (low entropic costs upon binding) and specificity (conformational restriction). p18, a natural AR protein interacting with CDK4/6 (**Fig. 5**), also has a low r.m.s.d.$_{C\alpha}$ of < 0.9 Å between the complexed (PDB entry 1G3N) and the uncomplexed (PDB entry 1IHB) AR protein. However, the interaction with CDK4/6 leads to structural alterations in the target[34]. In contrast to those of the rigid AR domain scaffolds, the binding site of antibodies seem to be able to adopt different conformations. The loops of the complementarity-determining region of typical protein binding antibodies can undergo substantial changes upon binding[35] and the loop flexibility might even be used to accomplish multispecificity[36].

Here, we have successfully validated our AR proteins as designed binding molecules using MBP and two MAPKs as model targets, and thus have introduced a binding molecule with very favorable properties. The combination of the high expression level of designed AR proteins (200 mg/l soluble protein in shake flasks), their high thermodynamic stability (9.5 to 21 kcal/mol[17]), the absence of cysteines, the fast enrichment of binders (four selection rounds), their low nanomolar affinities along with high specificities and their modular architecture compares favorably with reports on antibodies and other alternative scaffolds (such as protein A[28], lipocalins[14], fibronectin[37] or green fluorescent protein[38]). Our results open the door for a number of applications. Apart from being useful in capturing molecules on protein chips or in affinity purification—typical applications for designed binding molecules—designed AR proteins are especially suited for intracellular applications. Their high stability, the absence of disulfide bonds and the selectable high affinities are optimal prerequisites for intracellular inhibitors[39], where antibodies are less than ideal. Generally, AR proteins are also conceivable in therapeutic applications, which are currently a domain of recombinant antibodies. As we have shown here, it is possible to both cocrystallize a target protein with a designed binding AR protein and determine its crystal structure with the help of the AR protein. Hence, designed binding AR proteins could be used in cocrystallization and structure determination of proteins difficult to crystallize, similar to what has been shown with antibodies[40].

METHODS

Molecular biology. Unless stated otherwise, all experiments were done according to protocols found in reference 41. Enzymes and buffers were from New England Biolabs (NEB) or Fermentas. All PCR reactions were done using the proofreading Vent-polymerase (NEB).

Vectors used in antigen production. The different vectors that were prepared for the present study are described in detail in the **Supplementary Methods** online. pQEMBP (GenBank accession no. AY327141) was used for the expression of His-tagged, nonbiotinylated MBP. pAT224 (AY327139) was used for the expression of His-tagged, biotinylated MBP. pAT222 (AY327137) was used for the production of His-tagged, biotinylated pD. pAT222_JNK2 and pAT222_p38 were used for the production of His-tagged, biotinylated JNK2 and p38, respectively. All pAT222 and pAT224 constructs carry an Avi tag for biotinylation at the N terminus and a His_6 tag at the C terminus. pBirAcm (Avidity) was used for *in vivo* biotinylation.

Antigen production and purification. The biotinylated proteins pD, MBP, JNK2 and p38 (plasmids pAT222, pAT224, pAT222_JNK2 and pAT222_p38) were produced using *in vivo* biotinylation with plasmid pBirAcm in *E. coli* XL-1 Blue (Stratagene) according to the protocols of Avidity and QIAgen. Efficient biotinylation was confirmed by ELISA and blotting with a streptavidin-alkaline phosphatase conjugate (Roche) and mass spectrometry. Nonbiotinylated MBP for the ELISA analysis and crystallization was produced in the same way as the AR proteins[16] using pQEMBP in *E. coli* XL-1 Blue. The protein purification was carried out as described[16].

Ribosome-display vector (pRDV; AY327136). The cloning of pRDV is described in detail in the **Supplementary Methods** online. pRDV contains all flanking DNA regions necessary for ribosome display: the T7-promoter, the ribosomal binding site and an in-frame *tolA* gene spacer. Hence, by simple ligation of the DNA encoding the combinatorial library into pRDV and by a PCR using this ligation mix as template, all features necessary for ribosome display are added to the library. The use of pRDV has the advantage that it always provides error-free library flanking regions and that it saves a number of working steps compared to the standard PCR approach for library generation[21].

Generation of combinatorial libraries. The AR protein library generation has been described[16]. We changed that protocol in this study in that all ARs, that is,

the N-terminal capping AR, the designed repeat module and the C-terminal capping AR were used as PCR products for the assembly of the libraries. In the present study, both the N- and the C-terminal capping repeat were amplified by PCR from cloned and verified sequences to reduce sequence errors. In this way, AR protein libraries consisting of an N-terminal capping AR, two or three designed AR modules and a C-terminal capping AR (N2C and N3C libraries) were assembled. To convert the libraries to the ribosome-display format, they were amplified by PCR using oligonucleotides EWT4 (5′-TTCCTCATGAGAG GATCGCATCACCATCACCATCACGGATCCGACCTGGG-3′) and WTC4 (5′-TTTGGGAAGCTTTTGCAGGATTTCAGC-3′) and ligated into pRDV using the restriction enzymes *Bsp*HI (or *Nco*I for pRDV) and *Hin*dIII. The ligation product was purified using QIAquick (QIAgen) columns. The purified ligation served as template for a PCR using oligonucleotides RDVf1 (5′-CCTTTTGCT CACATGACCCG-3′) and tolAk (5′-CCGCACACCAGTAAGGTGTGCGGTT TCAGTTGCCGCTTTCTTTCT-3′). Thereby, combinatorial N2C and N3C DNA libraries were generated in the ribosome-display format.

Ribosome display. The PCR-amplified libraries were transcribed and selections were done as described[21]. For the selection, the biotinylated antigen was immobilized as follows: neutravidin (66 nM, 100 µl/well; Pierce) in TBS150 (50 mM Tris HCl, pH 7.4, 150 mM NaCl) was immobilized on a Maxisorp plate (Nunc) by overnight incubation at 4 °C. The wells were then blocked with 300 µl 0.5% BSA (Fluka) in TBS150 for 1 h at 23 °C. Biotinylated antigen (100 µl, 1 µM) in TBS150 with 0.5% BSA was allowed to bind for 1 h at 4 °C. Before the ribosome-display round, the wells were extensively washed with washing buffer WBT (50 mM Tris acetic acid, pH 7.5, 150 mM NaCl, 50 mM $Mg(CH_3COO^-)_2$, 0.05% Tween 20). A ribosome-display round consisted of two 30-min prepanning steps on neutravidin and a 1 h binding step on the target protein. After washing, RNA purification and reverse transcription (with oligonucleotide tolAk), a first PCR was done using oligonucleotides T7B (5′-ATACGAAAT TAATACGACTCACTATAGGGAGACCACAACGG-3′) and tolAk. This RT-PCR product was purified on an agarose gel and reamplified in a second PCR using the same oligonucleotides. The second PCR product served as template for the next round of ribosome display. The number of RT-PCR cycles was reduced from 40 to 30 to 25 in the first three rounds to monitor the enrichment of binders. Binders were analyzed after four or five rounds.

Analysis of selected binders. From the selected DNA pools, the AR open reading frame was amplified by PCR and cloned into pQE30 (QIAgen) via *Bam*HI/*Hin*dIII (oligonucleotides: EWT3: 5′-TTCCGCGGATCCGAC-CTGGG-3′ and WTC4). The DNA sequences were determined using standard techniques. The sequences of the MBP binding proteins off7 (AY326424), MBP3_5 (AY326425) and MBP3_16 (AY326426) have been deposited in GenBank. The amino acid sequences of all sequenced clones are listed in **Supplementary Figure 1** online. For the ELISA screening, the crude extract of 0.6 ml protein expression cultures was used (expression according to QIAgen). The cell pellets were lysed with 50 µl B-Per (Pierce) and the lysates were mixed with 250 µl TBS500 (50 mM Tris HCl, pH 8.0, 500 mM NaCl) each. For quantitative ELISA, BIAcore, CD, analytical gel-filtration and crystallization, single, selected library members were produced on a liter scale and purified as described[16]. CD spectroscopy and analytical gel-filtration were done as described[16] using 15 µM protein in TBS150 (pH 7.4).

ELISA. Biotinylated antigens were immobilized on neutravidin-coated plates as described above. For the screening of the pools, 100 µl of the above crude extracts were applied to wells with or without immobilized antigen for 1 h at 4 °C. After extensive washing with TBS150, binding was detected with an anti-RGS-His antibody (QIAgen; detects only the RGS-His_6-tag of the AR protein, not the His_6-tag of the antigen), an anti-mouse-IgG-alkaline phosphatase conjugate (Pierce) and p-nitrophenylphosphate (Fluka). Quantitative ELISAs were done in the same manner, except purified protein was used (see **Fig. 3**). For competition ELISA, the purified AR protein off7 was incubated with varying amounts of free MBP before (4 °C, 100 min) and during the binding reaction (see **Fig. 3**).

Surface plasmon resonance (SPR). SPR was measured using a BIAcore 3000 instrument (BIAcore). The running buffer was 20 mM HEPES, pH 7.4, 150 mM NaCl and 0.005% Tween 20. A streptavidin SA chip (BIAcore) was used with 480 RU biotinylated MBP immobilized (440 RU JNK2 and 450 RU p38, respectively). The interactions were measured at a flow of 60 µl/min with 5 min buffer flow, 2 min injection of MBP-binding AR protein in varying concentrations (10 pM to 200 nM) and an off-rate measurement of 40 min with buffer flow. The signal of an uncoated reference cell was subtracted from the measurements. Inhibition BIAcore measurements gave results similar to that of the kinetic analyses (data not shown). The p38 and JNK2 binders were measured similarly, but with an injection time of 3 min. The kinetic data of the interaction were evaluated with a global fit using BIAevaluation 3.0 (BIAcore), Scrubber (BioLogic software) and Clamp[42].

Complex purification and crystallization. MBP and the selected MBP-binding AR proteins off7, mbp3_5 and mbp3_16 were produced as described above. The cell pellets of 1-liter bacterial culture of each MBP and off7 (or mbp3_5 or mbp3_16) were pooled and then lysed using an Emulsiflex C5 (Avestin) followed by additional sonication. The proteins were purified using an IMAC column as described[16], followed by a preparative Superdex-75 (Amersham Pharmacia) size exclusion chromatography step in 10 mM Tris HCl, pH 7.6 and 100 mM NaCl. For every protein mixture, the peak fraction with the smallest molecular weight containing both MBP and the AR protein in equimolar amounts, as determined by SDS-PAGE, was collected and used for crystallization. Light scattering of these fractions was measured as described[17]. It showed a monodisperse particle distribution for all three complexes. For the off7/MBP complex (61.4 kDa calculated mass for the 1:1 complex) an average radius of 3.6 nm, equivalent to a hydrated particle of 95 kDa, was estimated, which corresponds to a nonhydrated particle of 65 kDa. The off7/MBP complex crystallized readily and was further analyzed. The protein complex was concentrated to 26 mg/ml for crystallization. Initial crystallization screening was done in 96-well, sitting drop, square well crystallization plates (Greiner Bio-One). The reservoirs were filled with 100 µl reservoir solution using an 8-channel pipette from a 2 ml 96-well, deep-well block into the crystallization plate. Using an 8-channel pipette, 2 µl of reservoir solution were pipetted in the crystallization well and mixed with 2 µl of protein solution. The initial crystals were refined using standard techniques. The crystals used for data collection grew in about 2–3 weeks in 30% PEG 6000, 0.1 M Tris HCl pH 8–9, 100 mM NaCl in a hanging drop experiment with 500 µl reservoir, 2 µl protein solution mixed with 2 µl water and 2 µl reservoir solution. For data collection the crystals were soaked in the mother liquid with 10% ethylene glycol for about 30 s to 1 min and flash frozen in a cryostream at 100 K.

Data collection, reduction, structure solution and refinement. Data were collected at the European Synchrotron Radiation Facility beamline ID14-1. The data were processed using MOSFLM, SCALA and TRUNCATE[43]. The crystal belonged to space group $P2_1$, with a Matthews coefficient of $V_M = 2.1$ Å^3/Da, corresponding to an estimated water content of 39%.

The crystal structure was determined by molecular replacement using the program AmoRe[44], with the structure of the unselected N3C library member E3_5 (PDB entry 1MJ0[17]) as a search model. A conventional AmoRe protocol (rotation, translation, rigid body refinement) was applied yielding a solution. This information was used to obtain a first electron density. At this point only the AR protein was clearly visible in the electron density. A solvent flipping protocol was then applied to modify the map[45]. MBP in its open form (PDB entry 1LLS)[20] was positioned in the resulting electron density using program O[46]. Because about a third of MBP was visible, manual building would have been feasible as well. The rest of the model building was carried out using the program O[46], the structure refinement was done in CNS[45] resulting in a final model with an R-factor of 19.5 % and an R_{free}-factor of 24.9 % (**Supplementary Table 1** online).

Analysis of the complexes. The structural analysis of the complexes was done as suggested[25]. H-bonds and hydrophobic interactions were calculated with HBPLUS[47], LIGPLOT[48] and DIMPLOT[48] using the default settings. Other parameters were calculated using the protein-protein interaction server http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html[25] or CNS[45].

The atomic coordinates of the described complex were deposited in the PDB (PDB-ID: 1SVX).

Note: Supplementary information is available on the Nature Biotechnology website.

ARTICLES

ACKNOWLEDGMENTS

We thank the members of the Plückthun and Grütter laboratories for valuable discussions and David L. Zechel for the critical reading of the manuscript. H.K.B. was supported by a pre-doctoral fellowship of the Roche Research Foundation. M.T.S. was in receipt of a pre-doctoral scholarship from the Fonds der Chemischen Industrie and the Bundesministerium für Bildung und Forschung. This work was supported by the Swiss National Center of Competence in Research in structural biology and the Swiss Krebsliga.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 24 December 2003; accepted 14 January 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Groves, M.R. & Barford, D. Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9**, 383–389 (1999).
2. Andrade, M.A., Perez-Iratxeta, C. & Ponting, C.P. Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131 (2001).
3. Kobe, B. & Kajava, A.V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
4. Marcotte, E.M., Pellegrini, M., Yeates, T.O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
5. Ellis, J., Dodds, P. & Pryor, T. Structure, function and evolution of plant disease resistance genes. *Curr. Opin. Plant Biol.* **3**, 278–284 (2000).
6. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins: Struct. Funct. Genet.* **17**, 363–374 (1993).
7. Sedgwick, S.G. & Smerdon, S.J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316 (1999).
8. Suzuki, F. *et al.* Functional interactions of transcription factor human GA-binding protein subunits. *J. Biol. Chem.* **273**, 29302–29308 (1998).
9. Malek, S., Huxford, T. & Ghosh, G. I B functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF- B. *J. Biol. Chem.* **273**, 25427–25435 (1998).
10. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349**, 293–299 (1991).
11. Plückthun, A. *et al.* in *Antibody Engineering* (eds. McCafferty, J., Hoogenboom, H.J. & Chiswell, D.J.) 203–252 (IRL Press, Oxford, 1996).
12. Nygren, P.-Å. & Uhlén, M. Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* **7**, 463–469 (1997).
13. Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167–187 (2000).
14. Skerra, A. Imitating the humoral immune response. *Curr. Opin. Chem. Biol.* **7**, 683–693 (2003).
15. Forrer, P., Stumpp, M.T., Binz, H.K. & Plückthun, A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539**, 2–6 (2003).
16. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).
17. Kohl, A. *et al.* Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. USA* **100**, 1700–1705 (2003).
18. Mosavi, L.K., Minor, D.L. Jr. & Peng, Z.-y. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. USA* **99**, 16029–16034 (2002).
19. Main, E.R., Xiong, Y., Cocco, M.J., D'Andrea, L. & Regan, L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure (Camb)* **11**, 497–508 (2003).
20. Rubin, S.M., Lee, S.Y., Ruiz, E.J., Pines, A. & Wemmer, D.E. Detection and characterization of xenon-binding sites in proteins by ^{129}Xe NMR spectroscopy. *J. Mol. Biol.* **322**, 425–440 (2002).
21. Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942 (1997).
22. Yang, F. *et al.* Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat. Struct. Biol.* **7**, 230–237 (2000).
23. Hon, W.C. *et al.* Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* **89**, 887–895 (1997).
24. Forrer, P., Tamaskovic, R. & Jaussi, R. Enzyme-linked immunosorbent assay for the measurement of JNK, ERK and p38 kinase activities. *Biol. Chem.* **379**, 1101–1111 (1998).
25. Jones, S. & Thornton, J.M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 13–20 (1996).
26. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
27. Högbom, M., Eklund, M., Nygren, P.-Å. & Nordlund, P. Structural basis for recognition by an *in vitro* evolved affibody. *Proc. Natl. Acad. Sci. USA* **100**, 3191–3196 (2003).
28. Eklund, M., Axelsson, L., Uhlén, M. & Nygren, P.-Å. Anti-idiotypic protein domains selected from protein A-based affibody libraries. *Proteins: Struct. Funct. Genet.* **48**, 454–462 (2002).
29. Wahlberg, E. *et al.* An affibody in complex with a target protein: structure and coupled folding. *Proc. Natl. Acad. Sci. USA* **100**, 3185–3190 (2003).
30. Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287–1292 (2000).
31. Wodak, S.J. & Janin, J. Structural basis of macromolecular recognition. *Adv. Protein Chem.* **61**, 9–73 (2002).
32. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Plückthun, A. Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. USA* **98**, 75–80 (2001).
33. Mian, I.S., Bradwell, A.R. & Olson, A.J. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217**, 133–151 (1991).
34. Jeffrey, P.D., Tong, L. & Pavletich, N.P. Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev.* **14**, 3115–3125 (2000).
35. Sundberg, E.J. & Mariuzza, R.A. Molecular recognition in antibody-antigen complexes. *Adv. Protein Chem.* **61**, 119–160 (2002).
36. James, L.C., Roversi, P. & Tawfik, D.S. Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367 (2003).
37. Xu, L. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9**, 933–942 (2002).
38. Zeytun, A., Jeromin, A., Scalettar, B.A., Waldo, G.S. & Bradbury, A.R. Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies. *Nat. Biotechnol.* **21**, 1473–1479 (2003).
39. Cattaneo, A. & Biocca, S. The selection of intracellular antibodies. *Trends Biotechnol.* **17**, 115–121 (1999).
40. Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842–846 (1995).
41. Sambrook, J., Fritsch, E.F. & Maniatis, T. Molecular cloning: A laboratory manual, edn. 2 (Cold Spring Harbor Press, New York, 1989).
42. Myszka, D.G. & Morton, T.A. CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.* **23**, 149–150 (1998).
43. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Navaza, J. *Acta Crystallogr. A* **50**, 157–163 (1994).
45. Brünger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
46. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. McDonald, I.K. & Thornton, J.M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
48. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
49. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
50. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary information on methods:

Antigen cloning and production: The vector for the expression of His-tagged, non-biotinylated MBP, pQE_MBP, was constructed by inserting the PCR amplified MBP gene into pAT115, a pAT37 derivative[1], in which the *Nhe*I/*Xba*I fragment was replaced by the corresponding fragment of pQE60 (QIAgen, Germany) to remove the additional *Nco*I site. The MBP gene was amplified from pMAL-c2X (New England Biolabs) with oligonucleotides MTS23 (5'-GGGAATTCTCATGAAAACTGAAGAAGGTAAACTGG-3') and MTS24 (5'-CGGGATCCAGTCTGCGCGTCTTTCAGG-3'), and the product was cloned into pAT115 with the restriction enzymes *Bsp*HI (or *Nco*I for pAT115) and *Bam*HI.

The vectors for the expression of biotinylated, His-tagged pD (pAT222) or MBP (pAT224) were generated in several steps. The basis of the constructs is pAT194, which was itself generated in two steps. First, the *Nhe*I/*Xba*I fragment of pAT115 (contains the terminator as well as a silent chloramphenicol gene) was replaced by the sole terminator sequence of pQE30 (QIAgen), which was amplified by PCR using oligonucleotides MTS19 (5'-GGGAATTCGCTAGCAGTACTGCGATGAGTGGCAG-3') and MTS20 (5'-GCTCTAGAGCGGCGGATTTGTCC-3'), yielding pQc-pD. Second, the lacI^q gene of pZS4-Int1[2] was PCR amplified using oligonucleotides MTS21 (5'-GGGAATTCGCTAGCCTAGGGAAGGCGAAGCGGCATGC-3') and MTS22 (5'-GCTCTAGAGATTTCCCTCGACAATTCGC-3'), cut with *Nhe*I/*Xba*I and inserted into *Xba*I-cut pQc-pD, yielding pAT194. This ligation allows the insertion of the lacI^q fragment in both orientations. We chose the orientation opposite to the T5 expression cassette. A functional test using *E. coli* strain BL21(DE3) showed that the plasmid pAT194 provided functional lac repressor protein.

pAT194 was used to construct the biotinylation vectors. For the insertion of the N-terminal biotinylation tag[3], oligonucleotides MTS33 (5'-AATTCATTAAAGAGGAGAAATTAACTATGGCTGGTCTGAAC-3') and MTS34 (5'-TATCGTTCAGACCAGCCATAGTTAATTTCTCCTCTTTAATG-3') as well as MTS35 (5'-GATATCTTCGAAGCTCAGAAAATCGAATGGCACGAAGGTTC-3') and MTS36 (5'-CATGGAACCTTCGTGCCATTCGATTTTCTGAGCTTCGAAGA-3') were annealed separately, phosphorylated using T4 polynucleotide kinase and ligated into pAT194, previously cut with *Eco*RI/*Nco*I, yielding pAT215. A PCR was made with pAT215 as the template with oligonucleotides oli266 (5'-CGTAGCTCAGCTCATTAGTGATGGTGATGGTGATGAGAAGCTTGGGCTGCAGGTCGACCC-3') and oli267 (5'-ATCGGATCCATGGGCACCGCAACCGCGCCC-3') and the resulting PCR product was cloned into pAT215 via *Nco*I/*Bpu*1102I, yielding pAT222. pAT222 has an ORF under the control of a T5 promotor, consisting of the biotinylation tag:phage lambda protein D:multiple cloning site:His_6 tag. The phage lambda protein D in pAT222 has a truncated N-terminus. This was done to reduce the size of the flexible N-terminus in order to retain high level expression. pAT224, the vector for the production of biotinylated MBP, was generated on the basis of pAT222, replacing the protein D gene *Nco*I/*Bam*HI by a PCR amplified MBP gene of pQE-MBP (oligonucleotides MBPavitagf (5'-GAAGGTTCCATGGGGAAAACTGAAGAAGGTAAACTG-3') and MTS24).

The vectors for the expression of biotinylated, His-tagged JNK2 (pAT222_JNK2) and p38 (pAT222_p38) were generated by PCR cloning. The kinase templates[4] were amplified using oligonucleotides jnk2f (5'-TTCCGCGGATCCGGTACCTCCGACTCTAAATGTGACAGTCAG-3') and jnk2r (5'-AAACCCAAGCTTGTCGACAGCCTTCAAGG-3') or p38f (5'-TTCCGCGGATCCGGTACCTCTCAGGAGCGTCCGACCTTCTACCGTCAGGAGCTGAACAAG-3') and p38r2 (5'-AAACCCCTGCAGGGGACTCCATCTCTTCTTGGTC-3') and the PCR products were cloned into pAT222 via *Bam*HI/*Hin*dIII (JNK2) or *Bam*HI/*Pst*I (p38).

Protein production and purification: The biotinylated proteins pD, MBP, JNK2 and p38 (plasmids pAT222, pAT224, pAT222_JNK2 and pAT222_p38) were produced by *in vivo* biotinylation with plasmid pBirAcm of Avidity Inc. (Denver, Co, USA) in *E. coli* XL-1 Blue (Stratagene, USA) according to the protocols of Avidity and QIAgen. Non-biotinylated MBP for analysis and crystallization was produced in the same way as the AR proteins[5] using pQE-MBP in *E. coli* XL-1 Blue. The protein purification has been described[5]. Biotinylation was confirmed using ELISA and blotting with a streptavidin-alkaline phosphatase conjugate (Roche, Basel, Switzerland) and mass spectrometry.

Ribosome-display vector (pRDV; AY327136): To create a vector suitable for ribosome display, β-lactamase (as a control insert[6]) was
PCR amplified with oligonucleotides blaf (5'-TATCCATGGCGGACTACAAAGATGACGATGACAAAGGATCCCACCCAGAAACGCTG-3') and rBLA (5'-TTGCTTCTGAAGCTTTCCCCAATGCTTAATC-3'), cut with *Hin*dIII and ligated to the *Hin*dIII cut *tolA* gene[6], which was amplified with the oligonucleotides tolAfor (5'-TATACCCAAGCTTTATATGGCCTCGGGGGCCGAATTCGGATCTGGTGGCCAGAAGCAAGCTGAA-3') and tolArev (5'-GGAAGATCTCTACTACGGTTTGAAGTCCAATGG-3'). The resulting β-lactamase:tolA fusion construct was cut *Nco*I/*Bgl*II and cloned into *Nco*I/*Bam*HI cut pTFT74[7]. The sequence of the resulting plasmid pRDV was verified using standard techniques.

References:

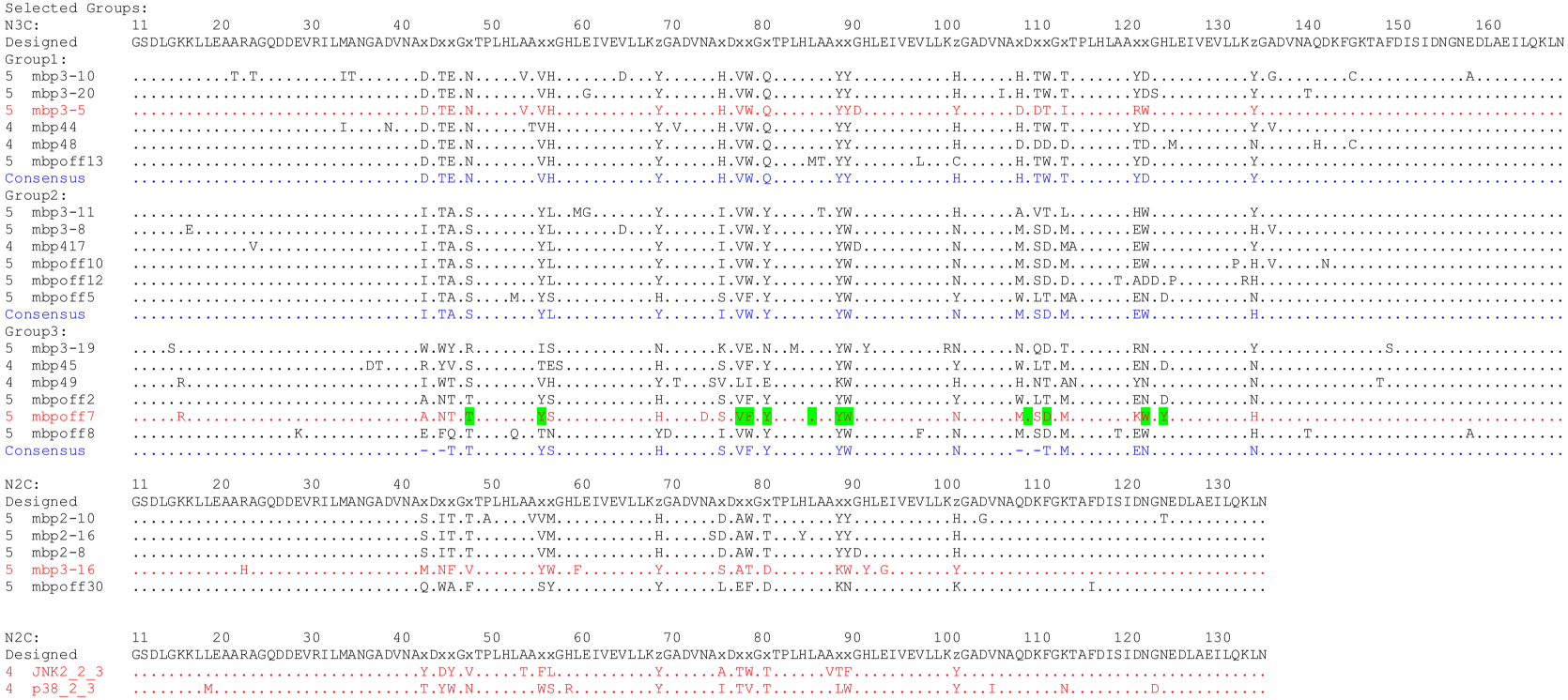
1. Forrer, P. & Jaussi, R. High-level expression of soluble heterologous proteins in the cytoplasm of *Escherichia coli* by fusion to the bacteriophage lambda head protein D. *Gene* **224**, 45-52 (1998).
2. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203-1210 (1997).
3. Cull, M.G. & Schatz, P.J. Biotinylation of proteins *in vivo* and *in vitro* using small peptide tags. *Methods Enzymol.* **326**, 430-440 (2000).
4. Forrer, P., Tamaskovic, R. & Jaussi, R. Enzyme-linked immunosorbent assay for the measurement of JNK, ERK and p38 kinase activity. *Biol. Chem.* **379**, 1101-1111 (1998).
5. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, woluble and wtable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489-503 (2003).
6. Amstutz, P. et al. *In vitro* selection for catalytic activity with ribosome display. *J. Am. Chem. Soc.* **124**, 9396-9403 (2002).
7. Ge, L., Lupas, A., Peraldi-Roux, S., Spada, S. & Plückthun, A. A mouse Ig kappa domain of very unusual framework structure loses function when converted to the consensus. *J. Biol. Chem.* **270**, 12446-12451 (1995).

Supplementary Figure 1

Sequences of designed AR proteins selected for MBP, JNK2 or p38 binding by ribosome display. The designed sequences for the N3C and the N2C libraries are given above the selected sequences (x represents a randomized potential interaction residue, where any amino acid was allowed except Cys, Gly or Pro; z represents a randomized framework residue where the three amino acids Asn, His or Tyr were allowed[1]). This restriction of allowed amino acids was achieved by mixed trinucleotides in the synthetic oligonucleotides used for AR assembly[2]. The residue numbers are given above the designed sequence. Only differences to the designed sequences are indicated. The names of the clones and the number of selection rounds are given on the left side of the respective sequence. The N3C sequences of the MBP binders are divided in three groups according to their sequences, and the respective group consensus sequences are shown below the sequences (in blue). The N2C sequences of the MBP binders as well as the JNK2 and p38 binders, respectively, are represented in two separate blocks. The five clones analyzed in more detail are highlighted in red font. The residues of off7 interacting with MBP are additionally highlighted in green. Note that in this representation, the N-terminal MRGSHHHHHH-tag (residues 1 to 10) has been removed.

References

1. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. & Plückthun, A. Designing Repeat Proteins: Well-Expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *J. Mol. Biol.* **332**, 489-503.
2. Virnekäs, B. *et al.* Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* **22**, 5600-5607 (1994).



Supplementary Table 1. Statistics for data collection and refinement of the off7/MBP complex.

Data collection:	
Space group	$P2_1$
Cell dimensions, Å	a= 74.68, b= 45.15, c= 79.57 α= γ= 90.0, β= 107.24
Resolution limits, Å	25.0 - 2.3
Observed reflections	total: 385166 unique: 22902
Completeness,%	98.7 (99.0)*
Redundancy	2.9
R_{sym} (% on I)	7.8 (34.8)*
Refinement:	
Resolution range, Å	2.3
R-factor/R_{free}, %	19.5/24.9
Ordered water molecules	214
rms deviation from ideal geometry:	
Bond lengths, Å	0.006
Bond angles, °	1.14
Average B factor, $Å^2$	34.6

* Numbers in parentheses refer to the highest-resolution shell.

Supplementary Table 2. List of the major hydrophobic interactions and H-bonds in the off7/MBP complex.

off7 interaction residue (repeat module #)[a]	H-bond[b]	MBP interaction residue (helix #)[a]
THR 48 *[c] (1)		**LYS 216 (H9)**
TYR 56 * (1)		**LYS 216 (H9)**
TYR 56 * (1)	OH<-NZ (3.11 Å)	**LYS 154 (H7)**
VAL 78 * (2)		**SER 366 (H17)**
VAL 78 * (2)		**GLY 367 (H17)**
PHE 79 * (2)		**LYS 214 (H9)**
PHE 79 * (2)		VAL 210 (H9)
PHE 79 * (2)		ALA 365 (H17)
TYR 81 * (2)		**LYS 214 (H9)**
TYR 81 * (2)	OH<-ND2 (2.80 Å)	**ASN 215 (H9)**
LEU 86 (2)		**LYS 216 (H9)**
TYR 89 * (2)		**ASN 215 (H9)**
TYR 89 * (2)		HIS 217 (H9)
TYR 89 * (2)	O<-NZ (2.85 Å)	**LYS 151 (H7)**
TRP 90 * (2)	NE1->O (3.08 Å)	**ASN 215 (H9)**
TRP 90 * (2)		HIS 217 (H9)
TRP 90 * (2)		ASP 150 (H7)
ASP 110 (3)	OD2<-NZ (2.68 Å)	**LYS 214 (H9)**
ASP 112 * (3)		**LYS 214 (H9)**
TRP 123 * (3)		**LYS 151 (H7)**
TRP 123 * (3)		ALA 148 (H7)
TYR 125 (3)	OH<-NZ (2.48 Å)	**LYS 151 (H7)**

[a] A cut-off of 3.9 Å was applied for hydrophobic interactions. Bold entries denote residues contributing with more than 5% to the total buried surface area.
[b] A cut-off of 3.9 Å for donor-acceptor and 2.5 Å for hydrogen-acceptor was applied.
[c] A star indicates that this amino acid is located in a randomized library position of off7.

Chapter 5

Kinase Inhibitors from Designed Ankyrin Repeat Protein Libraries: Selection and Analysis

Contents

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 280, No. 26, Issue of July 1, pp. 24715–24722, 2005

Printed in U.S.A.

Intracellular Kinase Inhibitors Selected from Combinatorial Libraries of Designed Ankyrin Repeat Proteins*[S]

Received for publication, February 15, 2005, and in revised form, April 25, 2005
Published, JBC Papers in Press, April 25, 2005, DOI 10.1074/jbc.M501746200

Patrick Amstutz‡, H. Kaspar Binz‡§, Petra Parizek, Michael T. Stumpp, Andreas Kohl, Markus G. Grütter, Patrik Forrer¶, and Andreas Plückthun‖

From the Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, Zürich CH-8057, Switzerland

The specific intracellular inhibition of protein activity at the protein level allows the determination of protein function in the cellular context. We demonstrate here the use of designed ankyrin repeat proteins as tailor-made intracellular kinase inhibitors. The target was aminoglycoside phosphotransferase (3′)-IIIa (APH), which mediates resistance to aminoglycoside antibiotics in pathogenic bacteria and shares structural homology with eukaryotic protein kinases. Combining a selection and screening approach, we isolated 198 potential APH inhibitors from highly diverse combinatorial libraries of designed ankyrin repeat proteins. A detailed analysis of several inhibitors revealed that they bind APH with high specificity and with affinities down to the subnanomolar range. *In vitro*, the most potent inhibitors showed complete enzyme inhibition, and *in vivo*, a phenotype comparable with the gene knockout was observed, fully restoring antibiotic sensitivity in resistant bacteria. These results underline the great potential of designed ankyrin repeat proteins for modulation of intracellular protein function.

In the postgenomic era, the functional analysis of the proteins encoded in the genomes is the rate-limiting step in understanding cellular processes. Most approaches in use today do not target the protein of interest directly but prevent its synthesis, either by total gene knockout (*e.g.* by homologous recombination), by antisense technologies, or by RNA interference (RNAi)[1] (1, 2). Although these technologies have generated remarkable insights into some intracellular protein functions, they suffer from a number of limitations. Total gene knockouts are often difficult to achieve and might even be lethal to the studied organism. Furthermore, the effects can be compromised by compensating changes in the regulation of other genes. The effect of antisense technologies is often not very strong. In a number of cases, even an 80% reduction of mRNA has no effect on the phenotype caused by the protein (3). Furthermore, the interpretation of results obtained from RNAi can be complicated by the up-regulation of other genes, limitation in specificity, and induction of the interferon response (4, 5). Finally, all methods that prevent the synthesis of the entire protein only allow the description of a loss-of-function phenotype.

The use of molecules acting at the protein level is a complementary approach. This approach has the additional advantage that single functions, domains or posttranslational modifications of a protein may be targeted, without removing the whole protein. Inhibitors of proteins can also be used in subsequent studies of the target protein *in vitro* and might then even be used as first lead compounds in the drug discovery process or at least simulate the effect of a future small molecule drug. Both constrained peptides (6–9) and single-chain antibody fragments (scFv (10)) have been applied for *in vivo* protein inhibition. In most cases, it is rather difficult to obtain high-affinity peptides to a target protein due to their high flexibility and the resulting loss of peptide chain entropy upon binding, limiting the free energy of binding, and thus the generation of high-affinity antibody fragments to a given target has been pursued more often (11). However, the drawback of antibodies is that their functionality is compromised in the reducing environment inside the cell, since the stabilizing intramolecular disulfide bonds are not formed (12). Whereas a number of antibody frameworks have been reported that are stable in the reducing environment (10, 13–15), the stability in the reducing state still does not reach that of other proteins. The solution to this problem is to use alternative binding proteins without disulfide bonds. Such alternative scaffolds have been developed (16–18), but only one intracellular selection for binding molecules with the yeast two-hybrid system has been published (19); no intracellular inhibitors have been reported to date.

We previously reported the generation of designed ankyrin repeat (AR) proteins as high-affinity binders with a great potential for intracellular applications (20, 21). Next to antibodies, repeat proteins are the most abundant class of binding molecules found in nature (22–24). They often act as intracellular enzyme modulators (*e.g.* as inhibitors of kinases) (25). Unlike antibodies, AR proteins do not contain disulfide bonds and are found in all cellular compartments, mediating protein-protein interactions with affinities in the nanomolar to picomolar range. AR proteins are built of consecutive homologous repeats, which assemble to form elongated domains with a continuous hydrophobic core. The number of repeats found in natural AR proteins varies from 2 to over 30 (26, 27). We used

* This work was supported by the Swiss National Centre of Competence in Research in Structural Biology and the Swiss Krebsliga. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental data and two additional figures.

‡ These authors contributed equally to this work.
§ Recipient of a predoctoral fellowship of the Roche Research Foundation.
¶ To whom correspondence may be addressed. Tel.: 41-44-635-55-03; Fax: 41-44-635-57-12; E-mail: pforrer@bioc.unizh.ch.
‖ To whom correspondence may be addressed. Tel.: 41-44-635-55-70; Fax: 41-44-635-57-12; E-mail: plueckthun@bioc.unizh.ch.

[1] The abbreviations used are: RNAi, RNA interference; AR, ankyrin repeat; EPK, eukaryotic protein kinase; APH, aminoglycoside phosphotransferase (3′)-IIIa; IPTG, isopropyl 1-thio-β-D-galactopyranoside; MBP, maltose-binding protein; IMAC, immobilized metal-ion affinity chromatography; RT, reverse transcription; ELISA, enzyme-linked immunosorbent assay; BSA, bovine serum albumin; MIC, minimal inhibitory concentration.

this modular architecture to generate designed AR protein libraries. By structure and sequence consensus analyses, we engineered an AR module of 33 amino acids with fixed framework residues and randomized potential interaction residues (28). This module was used as building block to create combinatorial libraries of designed AR proteins of different repeat numbers (Fig. 1). Members of these libraries are very well expressed, monomeric in solution, highly soluble, and stable (28). By x-ray crystallography, the expected AR domain fold could be verified (29). Specific binders to maltose-binding protein (MBP) of *Escherichia coli* and two human mitogen-activated protein kinases (JNK2 and p38) have been selected with dissociation constants in the low nanomolar range (20).

Here, we report the first tailor-made enzyme inhibitors based on designed AR proteins. Using a combination of *in vitro* selection and *in vivo* screening, we isolated designed AR proteins that specifically bind and inhibit aminoglycoside phosphotransferase type (3′)-IIIa (APH). APH is a bacterial kinase conferring resistance to aminoglycoside antibiotics (30) and can serve as a model system, since it shows structural homology to eukaryotic protein kinases (EPKs). We show that the selected AR proteins inhibit APH both *in vitro* and *in vivo*. Based on these results, the use of designed AR proteins as specific kinase inhibitors and as general intracellularly active tools is discussed. We describe elsewhere (31) the crystal structure of one selected APH-inhibiting AR protein in complex with APH and report the details of the inhibition mechanism of this particular inhibitor.

MATERIALS AND METHODS

Molecular Biology—Unless stated otherwise, all experiments were performed according to the protocols of Ref. 32. Enzymes and buffers were from New England Biolabs (Beverly, MA) or Fermentas (Vilnius, Lithuania). Oligonucleotides were from Microsynth (Balgach, Switzerland).

Antigen Cloning and Production—The APH gene was amplified from the plasmid pETSACG1 (a kind gift from Prof. G. Wright (33, 34) with the oligonucleotides APHf (5′-GCCGGCCATGGGTACCGCTAAAATGAGAATA-3′) and APHr (5′-GCATAAGCTTTCATTAAAACAATTCATCCA-3′). The PCR product was purified (QIAquick; QIAgen, Hilden, Germany), cut with NcoI/HindIII, and ligated into pAT223 (GenBank™ accession number AY327138), which is a pQE30 (QIAgen, Hilden, Germany) derivative, yielding pAT223_APH. The expression from this vector yields a fusion protein with an N-terminal avi tag for *in vivo* biotinylation, phage λ protein D (pD) as fusion protein, a His_6 tag for immobilized metal ion affinity chromatography (IMAC) purification, and APH as target molecule (avi-pD-His_6-APH).

Protein Production and Purification—The biotinylated proteins avi-pD-His_6-APH and avi-pD-His_6 were produced according to protocols from Avidity (Denver, CO) and QIAgen by *in vivo* biotinylation in *E. coli* XL1-Blue (Stratagene, La Jolla, CA), which was co-transformed with pBirAcm (Avidity) and pAT223 or pAT223_APH, respectively. Successful biotinylation of the IMAC purified proteins was confirmed by detection of the biotin tag on immobilized proteins using ELISA and blotting procedures with a streptavidin-alkaline phosphatase conjugate (Roche Applied Science) as detection reagent.

Nonbiotinylated wild-type APH used for the *in vitro* kinetic assays was expressed from pETSACG1 (33, 34) in BL21 cells (Stratagene) and purified over a Q-Sepharose column (Amersham Biosciences), followed by gel filtration (Superdex 75; Amersham Biosciences) as described (35) with some modifications (for details, see Supplemental Data). The correct size of APH was confirmed by SDS-PAGE, mass spectrometry, and multi-angle static light scattering.

For protein production of the AR proteins, the corresponding genes were isolated from pQIA (see below) and inserted into pQE30 to avoid co-purification of APH at a later stage. The proteins were expressed and purified by IMAC as described (29).

In Vitro Selection with Ribosome Display—The PCR-amplified N2C and N3C AR protein libraries (20) were transcribed, and selections were performed by ribosome display as described (36) on avi-pD-His_6-APH immobilized over its biotinylated avi tag (for details, see Supplemental Data). The number of total reverse transcription (RT)-PCR cycles after each selection round was reduced from round to round from 40 to 35 to 30, adjusting to the yield due to enrichment of binders.

In Vivo Selection by Replica Plating—For the *in vivo* selection, the vector pQIA was constructed with β-lactamase and APH dicistronically arranged under the control of the β-lactamase promoter, as well as an expression cassette for AR proteins under the control of an IPTG-inducible T5-promoter (Fig. 3*A*). For details, see Supplemental Data. The pools of binders selected by ribosome display were cloned into pQIA using the restriction enzymes BamHI and HindIII, transformed into *E. coli* (XL1-Blue) and plated on LB agar plates, containing 1% glucose, 50 μg/ml ampicillin, and 100 μM IPTG, and incubated overnight at 37 °C. The colonies were replica-plated on selective plates containing 1% glucose, 50 μg/ml ampicillin, 500 μg/ml kanamycin A, and 100 μM IPTG using Accutran replica plater devices (Schleicher & Schuell), incubated for 2 h at 4 °C and finally incubated overnight at 37 °C. Colonies whose replica failed to grow on the selective plates were taken up in LB and replica-spotted again on selective plates to make sure no false positives were selected due to incomplete replica plating. By this procedure, 198 kanamycin A-sensitive colonies were identified. All of these clones were ranked for their kanamycin A sensitivity by spotting them on selection plates with kanamycin A at concentrations of 50, 100, 250, and 500 μg/ml. The plasmids of the most potent inhibitors were isolated and sequenced using standard DNA sequencing.

Determining the Minimal Inhibitory Concentrations (MICs)—Cultures of *E. coli* (XL1-Blue) harboring pQIA, encoding an AR protein, were grown in LB with 1% glucose and 50 μg/ml ampicillin (LB-Glc-ampicillin) at 37 °C overnight. The cultures were diluted 1:30 into 5 ml of the same medium and shaken at 37 °C until an A_{600} of 0.3 was reached. The cultures were then diluted 1:1000 (~2.5×10^5 colony-forming units/ml) into deep well plates (Abgene, Epsom, Surrey, UK) containing 1 ml of the different selection media: LB-Glc-ampicillin-IPTG (100 μM) plus kanamycin A (0, 1, 5, 10, 25, 50, 100, 150, 250, 500 or 1000 μg/ml) and as a control LB-Glc-ampicillin with 1000 μg/ml kanamycin A (without IPTG) and shaken at 37 °C for 24 h. Growth was monitored by following the A_{600}. MIC end points were determined to be the concentrations of aminoglycoside antibiotic where no visible growth was obtained.

Analytical Size Exclusion Chromatography—Gel filtration experiments were done on a SMART system (Amersham Biosciences) using a Superdex 75 PC 3.2/30 column or a Superdex 200 PC 3.2/30 column (Amersham Biosciences) at 20 or 37 °C. The void volume was experimentally determined by using blue dextran. Samples of AR_1, AR_3a, and AR_B (20 μl) were loaded at a concentration of 25 μM on the Superdex 200 column and were run at 60 μl/min in TBS_{150}, pH 7.5 (50 mM Tris HCl pH 7.5, 150 mM NaCl), at 20 °C. Samples of APH and APH in complex with the AR protein AR_3a (20 μl) were loaded at a concentration of 10 μM on the Superdex 75 column and were run at 60 μl/min in $HEPES_{200}$ (50 mM HEPES, pH 7.5, 100 mM NaCl, 40 mM KCl, 60 mM $MgCl_2$) at 37 °C. Cytochrome *c* (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), and β-amylase (200 kDa) were used as molecular mass standards.

ELISA—To determine the specificity of the interaction between the selected AR proteins and APH, ELISA experiments were performed as described (13). Biotinylated proteins (avi-pD-His_6-APH, avi-pD-His_6, avi-pD-JNK2-His_6, and avi-pD-p38-His_6, final concentration 100 nM) were immobilized in neutravidin-coated ELISA plates blocked with bovine serum albumin (BSA). AR proteins (200 nM) were added for 30 min, and subsequently, the plates were washed five times. A first antibody against the RGS-His_6 tag of the AR protein (QIAgen; detects only the RGS-His_6 tag of the AR protein but not the His_6 tag of the antigen), followed by a secondary antibody/alkaline phosphatase conjugate (Pierce), was used for detection. After the addition of substrate (*p*-nitrophenylphosphate; Fluka) and incubation at 37 °C for 45 min, the A_{405} was measured in a plate reader. All incubation steps, including the antibody or AR protein incubations, were carried out in TBS_{150} with 0.5% BSA and 0.05% Tween 20, whereas washing was done with TBS_{150} containing only Tween 20.

Surface Plasmon Resonance—SPR was measured using a BIAcore3000 instrument (BIAcore, Uppsala, Sweden) at 20 °C. The running buffer was 20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.005% Tween 20. An SA chip (BIAcore) was used with 600 RU biotinylated avi-pD-His_6-APH immobilized. The interactions were measured at a flow of 60 μl/min with 5-min buffer flow, 2-min injection of APH-binding AR protein in varying concentrations (1–200 nM), and an off-rate measurement of 40 min with buffer flow. The signal of an uncoated reference cell was subtracted from the measurements. The kinetic data of the interaction were evaluated using BIAevaluation 3.0 (BIAcore), and global fits were used to determine K_D. The affinities of clones AR_3a and AR_3b were confirmed by competition BIAcore (Supplemental Data).

In Vitro Inhibition Studies—APH activity assays were performed essentially as described (35) with small modifications to facilitate inhibition studies. Briefly, the kinase buffer we used (50 mM HEPES, pH 7.5, 160 mM NaCl, 40 mM KCl, 10 mM $MgCl_2$, 0.025% BSA, 5 units of pyruvate kinase/lactate dehydrogenase, 2.5 mM phosphoenolpyruvate, 0.12 mg/ml NADH, 1 mM ATP) contains HEPES instead of Tris and additionally 160 mM NaCl, 0.025% BSA, and a lower concentration of NADH (0.12 instead of 0.5 mg/ml). APH was present at a concentration of 200 nM, whereas the AR proteins were present at 10 μM for the kanamycin A studies. This mixture (990 μl) was preincubated for 15 min, and the assay was initiated by the addition of 10 μl of kanamycin A solution (K4000; 10 μM final concentration; Sigma). Both the preincubation and the assay were carried out at 37 °C. For the studies with amikacin, both the preincubation and the assay were carried out at 20 °C in the presence of 2 μM AR proteins, and the reaction was started with 10 μl of amikacin solution (A1774; 740 μM final concentration; Sigma). The initial rates were normalized for background activity by subtracting the activity monitored in the absence of substrate.

RESULTS

To identify APH inhibitors, we applied a two-step approach of *in vitro* selection and *in vivo* screening. First, we used ribosome display to select APH binders from highly diverse combinatorial AR protein libraries. The resulting pools of binders were subsequently screened for inhibitors by replica plating. As APH mediates bacterial resistance to aminoglycoside antibiotics, inhibitors of the enzyme give rise to a kanamycin-sensitive phenotype.

Selection for APH Binders with Ribosome Display—The AR protein libraries used here (20) were termed N2C and N3C. The digit describes the number of designed and randomized AR modules between the N- and C-terminal capping repeats in the library members (Fig. 1). Three rounds of ribosome display selection were carried out for both the N2C and the N3C AR protein libraries in parallel. Selections were performed using purified target protein. For this purpose, APH was expressed as a fusion to bacteriophage λ pD (37) with an N-terminal avi tag for *in vivo* biotinylation (38) and a His_6 tag in the linker for IMAC purification (avi-pD-His_6-APH). This target protein and the same construct without the kinase (avi-pD-His_6) were expressed in soluble and biotinylated form in *E. coli* by using the expression plasmids pAT223_APH and pAT223, respectively. For ribosome-display selections, purified biotinylated APH was immobilized on neutravidin-coated wells blocked with BSA. To eliminate BSA-, neutravidin-, and avi-pD-His_6-binding library members, a prepanning step on correspondingly immobilized avi-pD-His_6 was applied, where neutravidin and BSA were also present. In the actual panning step, binders to APH were thus selected. The relative amounts of ribosomal complexes binding to the target protein can be estimated by monitoring the yield from RT-PCR on agarose gels. We observed a significant increase of RT-PCR yield when comparing the results from the first and the second selection round (Fig. 2). To our knowledge, such pronounced enrichment after only one round of ribosome display selection has not been reported for any kind of naive library (scFv, cDNA, or random sequence). This might indicate that AR protein libraries and ribosome display are a very favorable combination. In the third selection round, the selected pools of binders were analyzed for their specificity, comparing the RT-PCR yields after selection against APH or the unrelated MBP (Fig. 2). High specificity to APH could be monitored by comparing the agarose gel band intensities of these RT-PCR products (Fig. 2). In summary, we observed a rapid enrichment of binders and generated pools of specific binders in only three rounds of ribosome-display selection.

Screening for APH Inhibitors by Replica Plating—The pools of specific binders were subjected to an *in vivo* screen, based on replica plating, to identify those APH-binding AR proteins that are also inhibitors. For this purpose, the selected pools of binders were cloned into the selection vector pQIA (Fig. 3*A*). This vector carries an expression cassette for the selected APH binders under the control of an IPTG-inducible T5 promoter.

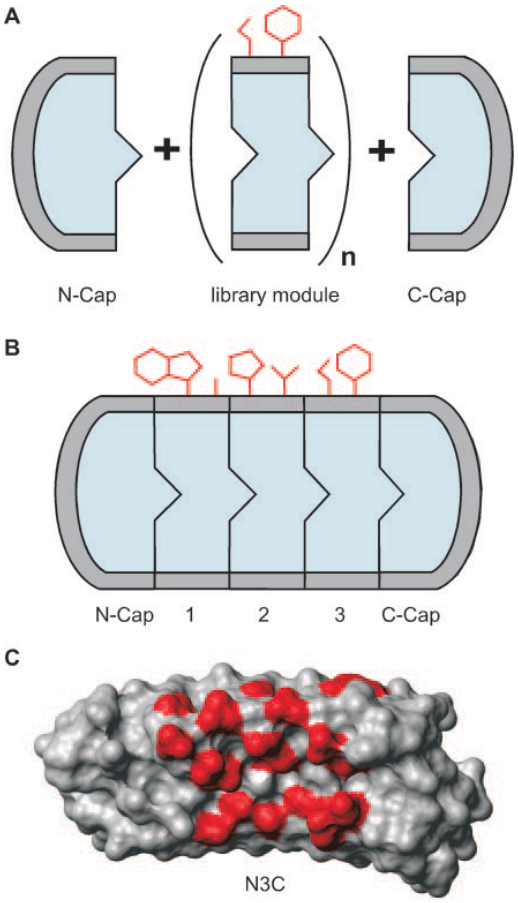


FIG. 1. **Designed AR protein libraries.** *A*, AR protein libraries are made from building blocks of designed self-compatible library modules with randomized positions (*red*), for potential target interaction and capping modules, which shield the hydrophobic core (*blue*). A varying number of library modules are genetically fused between an N- and a C-terminal capping repeat (*N-cap* and *C-cap*), yielding AR protein libraries of a specific size. *B*, an N3C library consists of proteins made from three random library modules, flanked by an N- and a C-terminal capping repeat. *C*, space-filling representation of an unselected N3C library member. The constant regions are *colored* in *gray*, and the library positions are depicted in *red*. The large potential interaction surface spans one side of the molecule (Protein Data Bank code 1MJ0) (29). For further details, see Refs. 21 and 28.

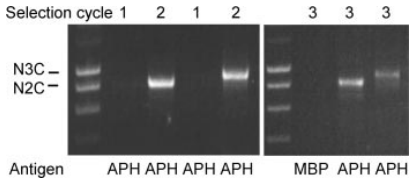


FIG. 2. **Enrichment and specificity of ribosome display selection rounds.** The outcome of ribosome-display selection rounds was monitored at the level of RT-PCR product yield by agarose gel electrophoresis. The RT-PCR yield of the first and the second selection round were compared for both the N2C and N3C library. In the third round, the specificities of the selected pools of the N2C and the N3C library were checked by panning against APH and an unrelated protein, MBP. All experiments were done in duplicates, which were combined for the agarose gel. For the specificity check of panning on the control protein MBP, both N2C and N3C duplicates were combined.

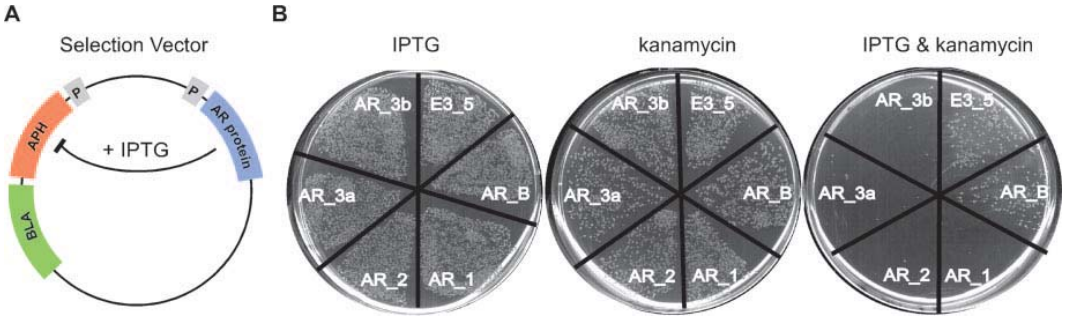


FIG. 3. **Phenotype of the selected clones.** *A*, the selection vector pQIA, carrying the β-lactamase (*BLA*) and APH gene for ampicillin and kanamycin resistance under a constitutive promoter in a dicistronic arrangement and the IPTG-inducible AR protein gene. Induction of the AR protein can result in inhibition of APH, if the AR protein has inhibitory properties, yielding a kanamycin-sensitive phenotype. *B*, the plating results of four APH inhibitors (AR_1, AR_2, AR_3a, AR_3b), one APH binder (AR_B), and an unrelated AR protein (E3_5) are shown. *E. coli* XL1-Blue cells were transformed with pQIA, carrying the respective genes in the AR protein cassette and plated under different conditions: nonselective in the case of only IPTG (100 μM), where the AR protein is expressed but no kanamycin A is present or if only kanamycin A is present (200 μg/ml) but the AR protein is not expressed; selective where the combination of IPTG and kanamycin A creates an environment where inhibitors repress growth.

pQIA also carries β-lactamase and APH as dicistronic genes under the control of the constitutively active β-lactamase promoter. Hence, this vector confers antibiotic resistance to ampicillin and kanamycin, of which the kanamycin resistance can be reverted, if an APH-inhibiting AR protein is expressed upon IPTG induction (Fig. 3*A*). As a control, we tested *E. coli* transformed with a pQIA variant harboring the unselected AR protein E3_5 (29) in the expression cassette. These cells were able to grow on plates with kanamycin A at concentrations of >1000 μg/ml, regardless of the presence of IPTG.

The pools of binders were cloned into pQIA and plated under nonselective conditions (ampicillin, IPTG, no kanamycin A). About 1000 colonies from the N3C pool and 2000 colonies from the N2C pool were subsequently replica-plated under selective conditions (ampicillin, IPTG, kanamycin A (500 μg/ml)). By this procedure, 129 N3C and 69 N2C clones were identified that showed a kanamycin-sensitive phenotype.

The identified clones were ranked for their *in vivo* inhibition efficiency by spotting them on selective plates with different kanamycin A concentrations, ranging from 10 to 500 μg/ml, and the most potent clones were sequenced. The sequencing revealed that variants of one potent inhibitor (AR_3 family) dominated the N3C inhibitor population, but also other sequences had been selected. Four different APH inhibitors (named AR_1, AR_2, AR_3a, and AR_3b), with different *in vivo* performances, were chosen for further characterization (for sequences see Supplemental Data). Whereas AR_1 was selected from the N2C library, AR_2, AR_3a, and AR_3b came from the N3C library. AR_3a and AR_3b differ by three point mutations at the DNA level, causing two amino acid mutations. We also studied the properties of clone AR_B, an N2C molecule that binds APH but fails to inhibit the target enzyme. As a further control, the unselected AR protein E3_5 (28, 29) was tested in parallel. The characterization of these AR proteins included expression tests, determination of minimal inhibitory concentrations (MICs), ELISA, BIAcore, and gel filtration studies as well as *in vitro* enzyme inhibition experiments.

In Vivo Performance (MIC)—The MICs of kanamycin A necessary to inhibit cell growth in liquid culture experiments were determined for the selected AR proteins. The plasmid pQIA with the nonbinding control AR protein E3_5 as insert mediated resistance to kanamycin A (MIC > 1000 μg/ml) regardless of the presence of IPTG, showing that the expression of an AR protein *per se* has neither an influence on kanamycin A resistance nor on bacterial growth. Clone AR_B showed the same phenotype as E3_5 (Fig. 3*B*). All AR proteins conferring kanamycin A sensitivity did not influence *E. coli* growth by themselves, even when induced, as long as no kanamycin A was present (Fig. 3*B*). In contrast, induction of the inhibitors stopped growth at different kanamycin A concentrations ranging from 10 to 150 μg/ml (Table I and Fig. 3*B*). Whereas clones AR_1 and AR_2 had MICs of 100 μg/ml and 150 μg/ml, respectively, clones AR_3a (25 μg/ml) and AR_3b (10 μg/ml) were more potent inhibitors (Table I). To put these results into perspective, we also evaluated the natural resistance of *E. coli* to kanamycin A in the absence of APH. The MIC of *E. coli* carrying a pQIA variant, where the APH gene had been removed, expressing AR_B or AR_3a was found to be 5 μg/ml each. Taken together, these results lead to the conclusion that the most potent clones AR_3a and AR_3b showed a phenotype comparable with the APH gene knockout.

Expression of the Selected AR Proteins—Since the selected AR proteins bind and inhibit APH inside the cell, we analyzed the expression of the selected clones in the selection vector pQIA under the expression conditions used for screening. Since the exact intracellular concentration of the AR proteins is difficult to determine, we performed a qualitative comparison of the selected clones by comparing the AR protein band intensities on Coomassie-stained SDS-PAGE. All AR proteins were expressed as soluble proteins regardless of the vector and the expression conditions (Fig. 4, *A* and *B*). The expression levels of clones AR_1, AR_3a, and AR_3b seem to be similar and around 2-fold higher than the one of clone AR_2. For *in vitro* studies, we recloned the AR proteins into an expression vector (pQE30) to avoid co-purification of APH. All AR proteins could be purified to homogeneity in a single IMAC step by means of their His_6 tag (Fig. 4*C*). Around 100 mg of purified AR protein was obtained from a standard 1-liter shake flask culture when expressed from the pQE30-derived vectors, without optimization of the expression and purification procedure. Even under selection conditions, where expression is not fully induced, the AR proteins show the most prominent bands on the gel (Fig. 4*B*). Since APH did not give rise to a pronounced band on the gel, the inhibitors must be present in large excess over APH in *E. coli* during screening.

Size Exclusion Chromatography of AR Proteins and Complexes—The selected AR proteins free or in complex with APH were analyzed by gel filtration experiments. Whereas most previously investigated AR proteins are monomers (20, 21), different oligomerization tendencies, depending on the individual protein, were detected for the AR proteins described here (Fig. 5*A*). AR_1 was monomeric, AR_B was present as monomer

TABLE I
Summary of ankyrin properties

Construct	MIC[a]	Expression level[b]	K_D[c]	k_{on}[c]	k_{off}[c]
	μg/ml		*nM*	$M^{-1}\,s^{-1}$	s^{-1}
No APH[d]	5				
E3_5[e]	>1000	+++			
AR_B[f]	>1000	+++	28.6 ± 16.2	$(1.4 \pm 0.5) \times 10^5$	$(4.1 \pm 0.8) \times 10^{-3}$
AR_1	100	+++	19.0 ± 10.0	$(1.2 \pm 0.4) \times 10^5$	$(2.2 \pm 0.4) \times 10^{-3}$
AR_2	150	+	8.1 ± 6.5	$(2.8 \pm 1.0) \times 10^5$	$(2.3 \pm 1.0) \times 10^{-3}$
AR_3a	25	+++	1.7 ± 1.3	$(1.6 \pm 0.6) \times 10^6$	$(2.7 \pm 0.9) \times 10^{-3}$
AR_3b	10	+++	0.5 ± 0.4	$(2.5 \pm 1.5) \times 10^6$	$(1.1 \pm 0.4) \times 10^{-3}$

[a] With kanamycin A.
[b] Estimated from corresponding band intensities from Coomassie-stained SDS-PAGE of whole cell extracts (Fig. 4*A*).
[c] Apparent values as described under "Materials and Methods," independently rounded.
[d] Control of *E. coli*, where no APH is present.
[e] Control of an unselected AR protein (28).
[f] Control of a selected binder, which did not show *in vivo* or *in vitro* inhibition.

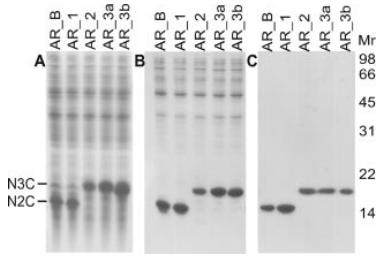


FIG. 4. **Expression and purification analysis of selected AR proteins.** The proteins were visualized on a 15% SDS-PAGE stained with Coomassie Brilliant Blue. *A*, expression of AR proteins under selection conditions: pQIA in *E. coli* XL1-Blue, carrying the respective genes in the AR protein cassette, induced with 100 μM IPTG. Cells were collected 4 h after induction and lysed in loading buffer. *B*, preparative expression of AR proteins with pQE30 (does not contain the APH gene) in *E. coli* XL1-Blue, induced with 500 μM IPTG. Cells were collected 4 h after induction and lysed in loading buffer. Crude extracts are shown. *C*, single step IMAC-purified AR proteins from the soluble *E. coli* fraction of the preparative expression. The size marker is indicated in kDa.

and trimer, and AR_2, AR_3a, and AR_3b appeared to be a mixture of monomer, dimer, and soluble multimer (see Fig. 5*A* for data of AR_B, AR_1, and AR_3a). Reapplying single peaks from AR_3a resulted in a similar elution pattern, indicating equilibrium between the oligomeric states. The exact molecular weight of the different species was verified for AR_1 and AR_B by static multi-angle light scattering (data not shown). Clone AR_1 (14.8 kDa) was determined to have an apparent molecular mass of 15.2 kDa, and clone AR_B (14.8 kDa) of 15.9 and 49.1 kDa. The molecular weight of the presumably monomeric and dimeric species of AR_3a and AR_3b could not be determined by static light scattering, since the peaks strongly overlapped. Since AR_1 and most other selected and unselected library members are monomers (20, 21), it seems likely that the oligomerization behavior is a consequence of the particular selected interaction residues of the individual AR proteins.

The complexes of the AR proteins with APH were also analyzed by gel filtration experiments. All complexes gave one distinct symmetric peak, which was shifted to higher molecular weight, compared with the one of APH alone, and which corresponded to the correct molecular weight of the complex. For example, the complex of APH and AR_3a eluted in a single peak at an apparent molecular mass of about 47 kDa (calculated molecular mass 49.6 kDa; Fig. 5*B*), although AR_3a alone eluted in various oligomeric states (Fig. 5*A*). The complexes showed no signs of oligomeric states, supporting the hypothesis that the selected binding interfaces of the AR proteins are responsible for and may be directly involved in the observed partial oligomerization of some AR proteins (Fig. 5*A*). Binding of the target thus abolishes oligomerization and leads to a defined 1:1 APH-AR protein complex. Indeed, co-purification of an APH mutant bound to AR_3a was used successfully for crystallization and structure determination of the complex as described elsewhere (31).

Binding Constants and Specificities of the Inhibitors—To investigate the binding properties of the selected AR proteins, the affinity and the specificity of the interaction were analyzed by ELISA and BIAcore studies. In an ELISA experiment with purified AR proteins of all analyzed clones, the binding to APH was compared with the binding to other proteins, such as pD, JNK2, p38, BSA, and neutravidin (Fig. 6). p38 and JNK2 are two EPKs (39) sharing structural homology to APH. The selected AR proteins bind specifically to APH but not to any of the other proteins (Fig. 6). As expected, the unselected AR protein E3_5 fails to bind any of the immobilized proteins. We conclude that specific APH-binding AR proteins were selected.

To determine the dissociation constants of the selected AR proteins to APH, kinetic surface plasmon resonance experiments were performed. The dissociation constants of all APH binders were found to be in the low nanomolar to subnanomolar range (Table I). K_D values of the two high-affinity inhibitors AR_3a and AR_3b were verified by competition BIAcore (40); for details, see Supplemental Data. The surface plasmon resonance data were fitted with the assumption that all species (monomer and dimer; see above) share the same binding properties or convert to monomers upon binding. If this is not the case, the K_D values given here reflect the average of all states.

In Vitro Inhibition of APH—To confirm that the *in vivo* effects of the selected AR proteins were indeed due to direct inhibition of the APH, *in vitro* enzyme assays were performed. Enzyme activity was monitored by coupling the release of ADP to a pyruvate kinase/lactate dehydrogenase reaction as described previously (35) with some modifications to adapt the assay for the inhibition studies (see "Materials and Methods"). The inhibition assay was set up to mimic the *in vivo* situation, where the AR protein concentration is higher than the one of APH and the kanamycin A concentration is low, since it has been suggested that kanamycin A import is the rate-limiting process (41). For all clones showing a kanamycin A-sensitive phenotype *in vivo*, enzyme inhibition was detected, whereas the control proteins E3_5 and AR_B showed almost no influence on APH activity (Fig. 7). Under these conditions (200 nM APH, 10 μM AR protein, kanamycin as substrate), the inhibition was not complete; a remaining activity, ranging from 6 to 52%, depending on the individual inhibitor, was found (Fig. 7*A*). The kinetic experiments with kanamycin A were, however, complicated by an apparent substrate inhibition at low micromolar concentrations (Fig. 7*C*). Substrate inhibition had previously been described at higher kanamycin A concen-

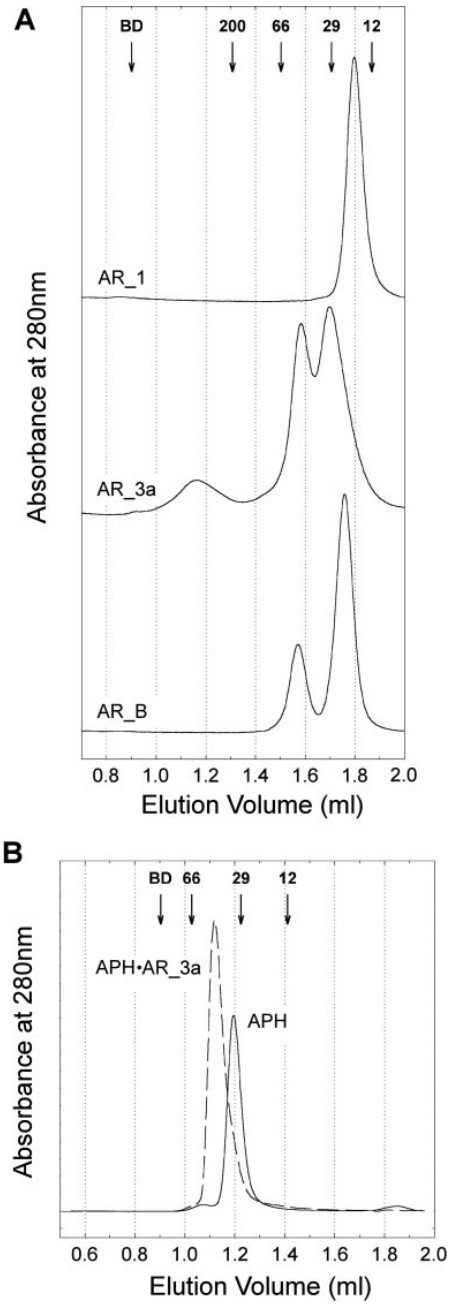


FIG. 5. **Size exclusion chromatography of selected AR proteins alone and in complex with APH.** The *arrows* indicate the elution volumes of the marker proteins (MW in kDa given above the arrow) and of blue dextran (*BD*). *A*, elution profiles of the AR protein clones AR_1, AR_3a, and AR_B from a Superdex 200 column. Clone AR_1 (14.8 kDa) eluted at an apparent molecular mass of 17 kDa; clone AR_3a (18.6 kDa) eluted at 400, 49, and 27 kDa; and clone AR_B (14.8 kDa) eluted at 52 and 20 kDa. *B*, elution profiles of APH and APH in complex with the AR protein AR_3a from a Superdex 75 column. The apparent molecular masses of the APH·AR_3a complex (49.5 kDa) and APH (30.9 kDa) are 47 and 34 kDa, respectively.

trations (35). The measurements were thus repeated with amikacin, another aminoglycoside substrate of APH containing an additional hydroxybutyrate group, which does not show substrate inhibition and has a similar turnover rate (35) (Fig. 7, *C* and *D*). For this substrate, APH activity was reduced to 2–73%,

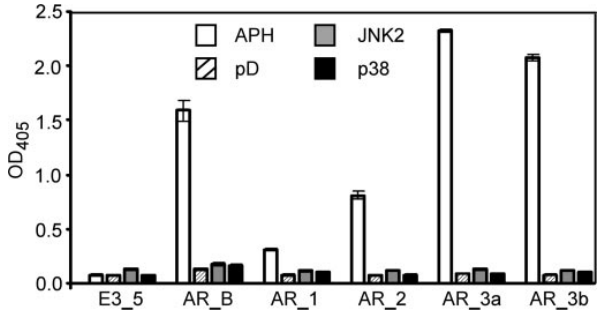


FIG. 6. **Binding specificities of the selected AR proteins.** The interaction of the selected AR proteins (AR_B, AR_1, AR_2, AR_3a, and AR_3b; all 200 nM) with immobilized APH was compared with the interaction with immobilized pD, p38, JNK2, neutravidin, and BSA. As a control, an unselected library member (E3_5; 200 nM) was included. The *error bars* show the standard deviation between three cells. The background binding of the detection antibodies was not subtracted.

depending on the respective AR inhibitor, whereas control proteins E3_5 and AR_B show almost no influence on activity (Fig. 7*B*). Thus, the overall *in vitro* inhibition of APH by the AR proteins using amikacin as substrate was similar to that observed when using kanamycin A (Fig. 7*A*), with almost complete inhibition for AR_3a and AR_3b (Fig. 7*B*). A more detailed characterization of AR_3a is given elsewhere, including precise titration analysis and structural insights in the inhibition mechanism (31).

The different inhibition efficiencies of the selected inhibitors seem to correlate well with the differences in K_D values and their inhibition performance *in vivo* (Table I). Nevertheless, it is worth commenting on the fact that residual activities were found at a concentration of AR proteins that is 10^3- to 10^4-fold above the K_D values determined by BIAcore analysis. Some of the APH·AR protein complexes may indeed possess a residual activity. However, there are other possible factors leading to this observation: (i) the presence of the substrates, ATP and aminoglycoside, acting as competitors for a distinct conformation of APH, which will raise K_I correspondingly above the K_D values; (ii) increased K_D values of the complexes as a result of different assay conditions of the kinetic measurements (*e.g.* increased temperature and different buffer composition); (iii) overestimation of the concentration of active AR proteins due to some degree of oligomerization (Fig. 5*A*).

One more difference should be noted: the inhibition of APH by AR_1 was less efficient with amikacin as a substrate (73% residual APH activity) than with kanamycin A (43% residual APH activity) under the assay conditions (Fig. 7, *A* and *B*). *In vivo* tests revealed that AR_1, in contrast to the other inhibitors, did not increase the amikacin sensitivity of *E. coli* harboring the appropriate plasmid but displayed the same phenotype as the noninhibitors E3_5 and AR_B (data not shown). These results suggest that amikacin can compete more efficiently than kanamycin A with the binding of AR_1 to APH.

In summary, all AR proteins tested showed direct inhibition of APH, although the inhibition was not complete except for AR_3b and AR_3a and might be substrate-specific in the case of AR_1.

DISCUSSION

Methods that are able to directly and specifically interfere with the function of a protein of interest intracellularly are necessary to validate results obtained from gene knockouts and especially from RNAi and antisense approaches. In addition, they give further insights into the target protein itself. The difficulty to create specific and intracellularly active binders and inhibitors to most proteins has, however, so far constituted

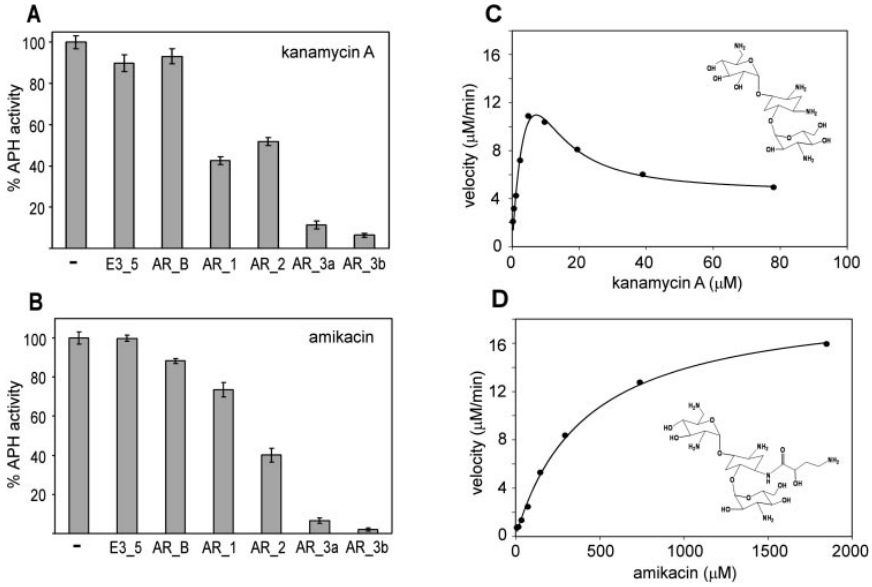


FIG. 7. **Inhibition properties of selected AR proteins and substrate-specific *in vitro* kinetics of APH.** *A* and *B*, the enzymatic activity of APH (200 nM) was determined with and without a large excess of AR protein present. *A*, with kanamycin A as substrate (10 μM) and AR proteins at a concentration of 10 μM, E3_5 and AR_B have no significant influence on APH activity, whereas the other AR proteins reduce the activity to 6–52%, depending on the inhibitor. *B*, with amikacin as substrate (740 μM) and in the presence of 2 μM AR proteins, the corresponding APH activity is reduced to 2–73%. Again, the control proteins E3_5 and AR_B show no significant influence on APH activity. For AR_3b, the inhibition is essentially complete. *C* and *D*, APH kinetics were measured at 37 °C using 50 nM purified monomeric APH, as described under "Materials and Methods," and two different substrates, kanamycin A (*C*) and amikacin (*D*). Initial velocities are plotted against substrate concentrations. *C*, when using kanamycin A as a substrate, a pronounced apparent partial substrate inhibition in the low micromolar range was observed. At high kanamycin A concentrations, k_{obs} reaches 1.64 s^{-1}. Although substrate inhibition of kanamycin A was previously reported (34), it seemed to be observed at lower substrate concentrations under our experimental conditions. *D*, the activity of APH followed Michaelis-Menten kinetics when using the substrate amikacin. The initial rates were fitted by nonlinear least squares fitting using the standard Michaelis-Menten equation. The determined K_m and k_{cat} values were 418 ± 29 μM and 6.56 ± 0.17 s^{-1}, respectively.

a bottleneck in elucidating the function of many proteins. By the rapid generation of selective high affinity intracellular enzyme inhibitors, we demonstrate here that designed AR proteins have the potential to solve this problem. The most potent inhibitors lead to a phenotype that is comparable with a gene knockout. This potency is achieved by a combination of factors; the selected AR proteins are free of disulfide bonds, they are able to fold well and are stable in the reducing intracellular environment, they are expressed in soluble form at high levels in the cytoplasm, and the interaction with the target enzyme is of high affinity and specificity. The combination of these factors leads to a clear advantage of AR proteins as intracellular inhibitors over antibodies, peptides, or other binding proteins reported so far.

Intracellular Protein Inhibitors—The approach to obtain intracellular protein inhibitors presented here, based on designed AR proteins, is very promising. Whereas high affinities are rather difficult to achieve with peptidic binders (except in the case of target proteins with pronounced binding pockets), the affinity and specificity of antibodies and designed AR proteins are good and comparable. The main drawback of antibodies for intracellular applications is that their stabilities rely on disulfide bonds, which cannot form in the reducing intracellular environment. Consequently, low levels of soluble functional protein and limited half-life of the antibody domains, due to aggregation, are the rule when applied in the cytoplasm (12). In this respect, the high intracellular expression level and stability of designed AR proteins is a clear advantage. It seems highly probable that these properties were responsible for the fact that we could identify not only a few individual but many different functional inhibitors, although the selection system described here selects for intracellular stability and enzyme inhibition at the same time. Several approaches have been described to generate antibodies that are active intracellularly (for details, see reviews (10, 15)). Although the inhibition of enzymes *in vitro* by antibodies has often been described (42–44), only a few of these inhibitors have been applied intracellularly (45).

Besides antibodies and AR proteins, other protein scaffolds have been used for the generation of enzyme inhibitors. Different high-affinity protease inhibitors have been generated based on protease-inhibitor scaffolds (54), and the function of β-lactamase inhibitory protein BLIP has been improved (46). These scaffolds mostly exploited naturally occurring enzyme inhibitors that were improved for affinity or specificity. All of these inhibitors contain disulfide bonds and were used for extracellular or periplasmatic applications.

In Vivo Inhibition of APH—The choice of APH as a model system was guided by several aspects of the enzyme. APH phosphorylates aminoglycoside antibiotics (*e.g.* streptomycin, kanamycin, or amikacin), mediating bacterial resistance to these antibiotics. This allowed the set-up of an efficient screening assay for the identification of inhibiting AR proteins (Fig. 3). Most importantly, APH displays high structural homology to EPKs (30), a class of enzymes involved in a wide range of diseases (see below). Finally, no satisfactory APH inhibitors are known to date that are active *in vivo*. Whereas different *in vitro* inhibitors of APH have been described in the literature (47, 48), they fail to inhibit bacterial growth efficiently. We describe here the first inhibitors that are fully functional *in vivo* and show a phenotype comparable with the APH gene knockout. Although the selected AR proteins are *per se* not suitable as antibiotic agents on their own, they could in the future accelerate the process of drug discovery by providing a detailed structure of the drug target in an inactive state. This

aspect is discussed elsewhere (31), where the crystal structure of the selected AR protein AR_3a in complex with an APH mutant is described.

Targeting Eukaryotic Protein Kinases—EPKs receive widespread interest, since they are key players in cellular signaling, making them targets for the treatment of a number of diseases, such as various cancers, asthma, and autoimmunity to name a few (49). In the case of EPKs, highly selective inhibitors are desired (for a recent review on the EPK inhibitor field see Ref. 50). However, the generation of specific EPK inhibitors is difficult, due to the large number of EPKs (about 518 in humans (49)), all sharing highly homologous active sites and using ATP as a co-factor. Most inhibition strategies involve targeting of the ATP binding pocket by small molecular weight drugs. This is not ideal, since this region of EPKs is highly conserved, and thus specificity is difficult to achieve and is further complicated by the high ATP level inside the cell (1 mM) that must be competed. Indeed, APH is inhibited by some EPK inhibitors, which on the one hand underlines the structural homology of APH and EPKs and on the other hand also demonstrates the poor specificity of these EPK inhibitors (51). Proteinaceous inhibitors could represent an attractive alternative, since they can interact via a large surface and, therefore, their binding is not restricted to grooves. This enables them to bind to the less conserved surface regions on the target protein while blocking its function.

It is noteworthy that many natural AR proteins are kinase inhibitors, such as for example the cell division kinase inhibitors $p16^{INK4a}$ and $p19^{INK4d}$ (52, 53). We have already successfully demonstrated the selection of mitogen-activated protein kinase binders from designed AR protein libraries (20). Again, the selected pools of binders might serve as starting points for inhibitor screens, in this case using an *in vitro* activity assay. Since mitogen-activated protein kinases are part of signaling networks, functional inhibition might not only be achieved by inhibition of the catalytic site, but also by simple binding, provided that the inhibitor blocks either the protein substrate or the activation of the kinase itself by an upstream kinase. Although our approach can be applied now as a discovery tool, therapeutic intracellular applications of designed AR proteins will have to await further maturation of gene therapy or protein uptake techniques.

Acknowledgments—We thank Christian Zahnd for valuable discussions and the groups of Prof. A. Plückthun and Prof. M. Grütter for an inspiring working atmosphere. We also thank David D. Boehr and Prof. Gerard D. Wright for help and discussion of APH kinetics.

REFERENCES

1. Dove, A. (2002) *Nat. Biotechnol.* **20,** 121–124
2. Dykxhoorn, D. M., Novina, C. D., and Sharp, P. A. (2003) *Nat. Rev. Mol. Cell Biol.* **4,** 457–467
3. Visintin, M., Meli, G. A., Cannistraci, I., and Cattaneo, A. (2004) *J. Immunol. Methods* **290,** 135–153
4. Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H., and Williams, B. R. (2003) *Nat. Cell. Biol.* **5**, 834–839
5. Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P. S. (2003) *Nat. Biotechnol.* **21**, 635–637
6. Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J., and Brent, R. (1996) *Nature* **380,** 548–550
7. Cohen, B., Ko Ferrigno, P., Silver, P. A., Brent, R., and Cohen, B. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13720–13725
8. Norman, T. C., Smith, D. L., Sorger, P. K., Drees, B. L., O'Rourke, S. M., Hughes, T. R., Roberts, C. J., Friend, S. H., Fields, S., and Murray, A. W. (1999) *Science* **285,** 591–595
9. Kamb, A., and Caponigro, G. (2001) *Curr. Opin. Chem. Biol.* **5,** 74–77
10. Cattaneo, A., and Biocca, S. (1999) *Trends Biotechnol.* **17,** 115–121
11. Hoogenboom, H. R., and Chames, P. (2000) *Immunol. Today* **21**, 371–378
12. Biocca, S., Ruberti, F., Tafani, M., Pierandrei-Amaldi, P., and Cattaneo, A. (1995) *Bio/Technology* **13**, 1110–1115
13. Wörn, A., and Plückthun, A. (1998) *FEBS Lett.* **427**, 357–361
14. Auf der Maur, A., Zahnd, C., Fischer, F., Spinelli, S., Honegger, A., Cambillau, C., Escher, D., Plückthun, A., and Barberis, A. (2002) *J. Biol. Chem.* **277**, 45075–45085
15. Lobato, M. N., and Rabbitts, T. H. (2003) *Trends Mol. Med.* **9,** 390–396
16. Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M., and Nygren, P.-Å. (1997) *Nat. Biotechnol.* **15,** 772–777
17. Koide, A., Bailey, C. W., Huang, X., and Koide, S. (1998) *J. Mol. Biol.* **284,** 1141–1151
18. Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W., and Lipovsek, D. (2002) *Chem. Biol.* **9,** 933–942
19. Koide, A., Abbatiello, S., Rothgery, L., and Koide, S. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1253–1258
20. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G., and Plückthun, A. (2004) *Nat. Biotechnol.* **22,** 575–582
21. Forrer, P., Stumpp, M. T., Binz, H. K., and Plückthun, A. (2003) *FEBS Lett.* **539,** 2–6
22. Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001) *J. Struct. Biol.* **134,** 117–131
23. Kobe, B., and Kajava, A. V. (2001) *Curr. Opin. Struct. Biol.* **11,** 725–732
24. Bork, P. (1993) *Proteins* **17,** 363–374
25. Chan, F. K., Zhang, J., Cheng, L., Shapiro, D. N., and Winoto, A. (1995) *Mol. Cell Biol.* **15,** 2682–2688
26. Sedgwick, S. G., and Smerdon, S. J. (1999) *Trends Biochem. Sci.* **24,** 311–316
27. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C., and Peng, Z. (2004) *Protein Sci.* **13,** 1435–1448
28. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., and Plückthun, A. (2003) *J. Mol. Biol.* **332,** 489–503
29. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A., and Grütter, M. G. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1700–1705
30. Hon, W. C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S., Wright, G. D., and Berghuis, A. M. (1997) *Cell* **89,** 887–895
31. Kohl, A., Amstutz, P., Parizek, P., Binz, H. K., Briand, C., Capitani, C., Forrer, P., Plückthun, A., and Grütter, M. G. (2005) *Structure,* in press
32. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Handbook*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York
33. Boehr, D. D., Thompson, P. R., and Wright, G. D. (2001) *J. Biol. Chem.* **276,** 23929–23936
34. Thompson, P. R., Hughes, D. W., and Wright, G. D. (1996) *Chem. Biol.* **3,** 747–755
35. McKay, G. A., Thompson, P. R., and Wright, G. D. (1994) *Biochemistry* **33,** 6936–6944
36. Hanes, J., and Plückthun, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4937–4942
37. Forrer, P., and Jaussi, R. (1998) *Gene* (*Amst.*) **224,** 45–52
38. Beckett, D., Kovaleva, E., and Schatz, P. J. (1999) *Protein Sci.* **8,** 921–929
39. Davis, R. J. (2000) *Cell* **103,** 239–252
40. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., and Plückthun, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14130–14135
41. McKay, G. A., and Wright, G. D. (1995) *J. Biol. Chem.* **270,** 24686–24692
42. Sun, J., Pons, J., and Craik, C. S. (2003) *Biochemistry* **42,** 892–900
43. Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S., and Cambillau, C. (2002) *J. Biol. Chem.* **277,** 23645–23650
44. Lauwereys, M., Arbabi Ghahroudi, M., Desmyter, A., Kinne, J., Holzer, W., De Genst, E., Wyns, L., and Muyldermans, S. (1998) *EMBO J.* **17,** 3512–3520
45. Jobling, S. A., Jarman, C., Teh, M. M., Holmberg, N., Blake, C., and Verhoeyen, M. E. (2003) *Nat. Biotechnol.* **21,** 77–80
46. Huang, W., Zhang, Z., and Palzkill, T. (2000) *J. Biol. Chem.* **275,** 14964–14968
47. Burk, D. L., and Berghuis, A. M. (2002) *Pharmacol. Ther.* **93,** 283–292
48. Kotra, L. P., Haddad, J., and Mobashery, S. (2000) *Antimicrob. Agents Chemother.* **44,** 3249–3256
49. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) *Science* **298,** 1912–1934
50. Noble, M. E., Endicott, J. A., and Johnson, L. N. (2004) *Science* **303,** 1800–1805
51. Daigle, D. M., McKay, G. A., and Wright, G. D. (1997) *J. Biol. Chem.* **272,** 24755–24758
52. Russo, A. A., Tong, L., Lee, J. O., Jeffrey, P. D., and Pavletich, N. P. (1998) *Nature* **395,** 237–243
53. Brotherton, D. H., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P. J., Volyanik, E., Xu, X., Parisini, E., Smith, B. O., Archer, S. J., Serrano, M., Brenner, S. L., Blundell, T. L., and Laue, E. D. (1998) *Nature* **395,** 244–250
54. Stoop, A. A., and Craik, C. S. (2003) *Nat. Biotechnol* **9,** 1063–1068

Supplemental Data

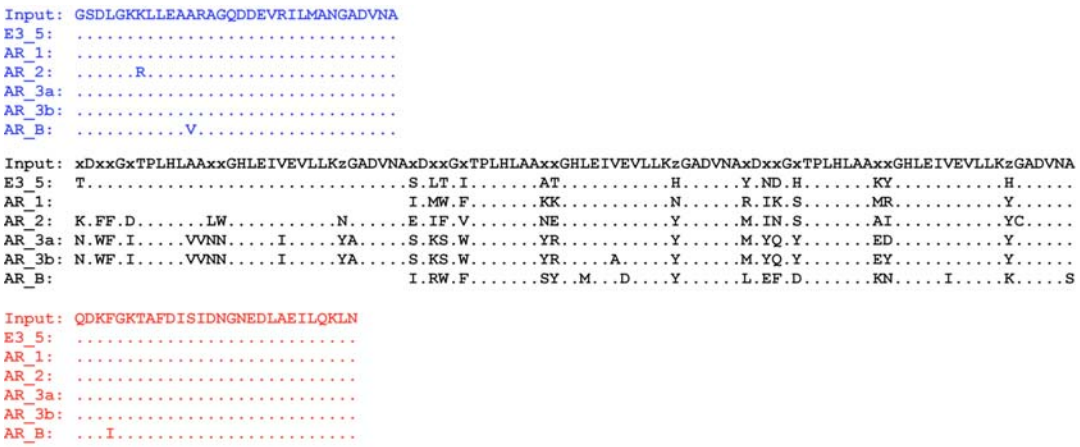
In vitro selection with ribosome display - The PCR amplified N2C and N3C AR protein libraries (13) were transcribed and selections were performed by ribosome display as described (29) on avi-pD-His_6-APH immobilized over its biotinylated avi-tag. For the selection, the biotinylated antigen was immobilized as follows: neutravidin (100 µl/well, 66 nM; PIERCE, Rockford, IL, USA) in TBS_{150} (50 mM Tris-HCl pH 7.4, 150 mM NaCl) was immobilized in a Maxisorp plate (Nunc; Roskilde, Denmark) overnight at 4°C. The wells were then blocked with 300 µl 0.5% BSA (Fluka; Buchs, Switzerland) in TBS_{150} for 1 h at 23°C. One hundred µl biotinylated antigen (concentration approximately 500 nM) in TBS_{150} with 0.5% BSA were then bound to the immobilized neutravidin for 1 h at 4°C. Prior to the ribosome display round, the wells were extensively washed with washing buffer containing Tween-20 (WBT; 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 50 mM $MgCl_2$, 0.05% Tween-20) (29). A standard round consisted of a 30-minute pre-panning step on neutravidin, another pre-panning step of 30 minutes on avi-pD-His_6, to avoid selecting binders against BSA, neutravidin or pD, followed by a 60-minute panning step on avi-pD-His_6-APH (bound to neutravidin, blocked with BSA). Following selection, RNA purification and RT (with oligonucleotide tolAk), a first PCR was made using oligonucleotides T7B (5'-ATACGAAATTAATACGACTCACTATAGGGAGACCACAACGG-3') and tolAk (5'-CCGCACACCAGTAAGGTGTGCGGTTTCAGTTGCCGCTTTCTTTCT-3'). This RT-PCR product was agarose-gel purified and reamplified in a second PCR using the same oligonucleotides. This second PCR product served as template for the next round of ribosome display. The number of total RT-PCR cycles was reduced from round to round from 40 to 35 to 30, adjusting to the yield due to enrichment of binders.

Construction of selection vector for replica screening - For the *in vivo* selection, the vector pQIA was constructed with □-lactamase and APH dicistronically arranged under the control of the □-lactamase promoter, as well as an expression cassette for AR proteins under the control of an IPTG-inducible T5-promoter (Fig. 3A). For this purpose, the APH gene was PCR amplified from pETSACG1 with the oligonucleotides APHsspf (5'-TTTCAATATTTTAAAACAATTCATCCAGTAAAATATAGTATTTTATTTTC-3') and APHsspr (5'-TTCAATAATATTGAAAAAGGAAGAGTATGGCTAAAATGAGAATAT-3'). The PCR product was purified, cut with *Ssp*I and ligated into pQI_pD (pAT194, GenBank accession number AY327142), a pQE30 derivative with additional $lacI^q$ gene and pD under the control of the T5 promoter, cut with the same enzyme, to yield pQIA_pD, where the APH open reading frame lies directly downstream of the one of □-lactamase in dicistronic manner. The proper function of the APH gene was confirmed by successful growth of XL1-Blue cells harboring pQIA_pD in the presence of 50 µg/ml kanamycin A. For later cloning of the APH-binding AR protein pool, the multiple cloning site and pD of pQIA_pD was replaced with the one of pPANK (GenBank accession number AY327140) containing the unselected AR protein E3_5 (GenBank accession number AY195853), yielding the *in vivo* selection plasmid pQIA containing the naïve AR protein E3_5 gene as a dummy insert under the control of the IPTG-inducible T5 promoter (Fig. 3A). The pools of binders selected by ribosome display were cloned into pQIA using the restriction enzymes *Bam*HI and *Hin*dIII, replacing the E3_5 gene, transformed into *E. coli* (XL1-Blue) and plated on LB agar plates, containing 1% glucose, 50 µg/ml ampicillin and 100 µM IPTG, and incubated overnight at 37°C.

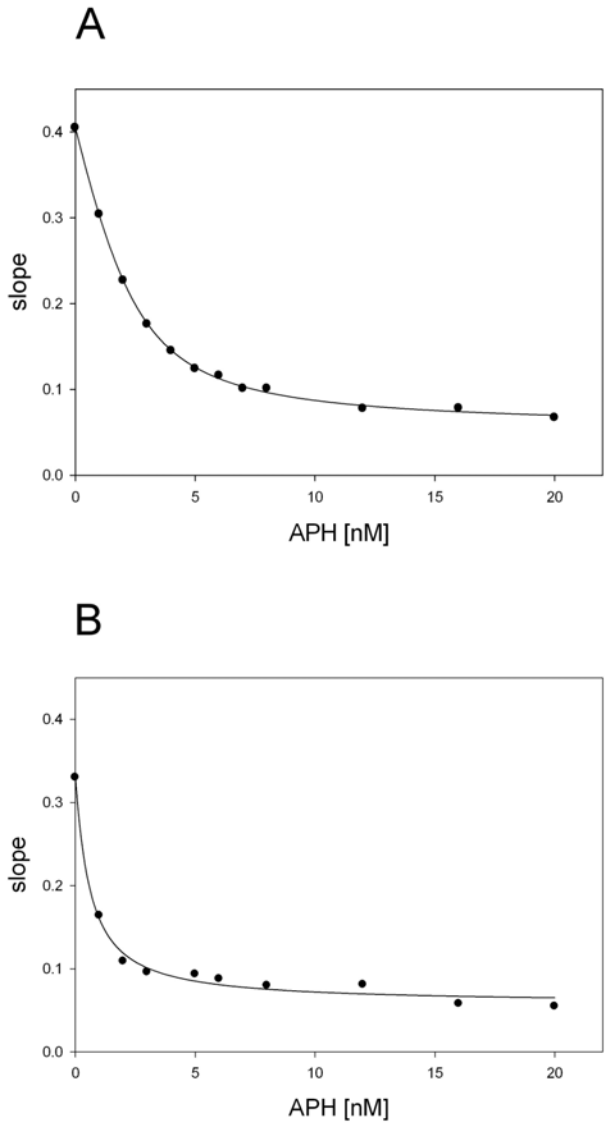
Protein production and purification - Non-biotinylated wild-type APH used for the *in vitro* kinetic assays was expressed from pETSACG1 in BL21 cells (Stratagene) and purified over a Q-Sepharose column (Amersham Pharmacia, Dübendorf, Switzerland), followed by gel-filtration (Superdex 75; Pharmacia) as described (28) with some modifications.
After expression, the cell pellet of one liter bacterial culture was resuspended in 25 ml TBS_{500} (50 mM Tris-HCl pH 8.0, 500 mM NaCl). The cells were lysed using an Emulsiflex C5 (Avestin, Ottawa, ON, Canada). The cell debris was pelleted by centrifugation at 30'000 g for 30 minutes at 4°C. The resulting supernatant was applied to a 25 ml Q-Sepharose column (Amersham Pharmacia, Dübendorf, Switzerland) equilibrated with buffer 1 (50 mM Tris-HCl pH 8.0). APH was eluted by a linear

gradient with increasing buffer 2 (50 mM Tris-HCl pH 8.0, 1 M NaCl). Fractions containing APH were identified by SDS-PAGE analysis, pooled, and 10 mM $MgCl_2$ and 2 mM dithiothreitol (DTT) were added. APH was further purified by a preparative Superdex-75 (Amersham Pharmacia, Dübendorf, Switzerland) size exclusion chromatography step in 50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$ and 2 mM DTT, according to the published protocol (Hon et al., 1997). The peak fraction with an approximate molecular weight of the monomeric APH was collected and used for further experiments. The correct size of APH was confirmed by SDS-PAGE, mass spectrometry and static light scattering.

Supplemental Figures



Supplemental Figure S1 - Sequences of the selected AR proteins. The N-terminal capping repeat is coloured in blue, the two or three randomized repeats are in black and the C-terminal capping repeat is in red. E3_5 is an unselected AR protein; AR_1, AR_2, AR_3a and AR_3b are APH inhibitors, while AR_B binds APH without inhibiting it. Note that AR_1 and AR_B are N2C molecules, while E3_5, AR_2, AR_3a and AR_3b are N3C molecules. For comparison, the N3C library input sequence is given on the top of the sequence alignment. In the alignment, only differences compared to the input sequence are highlighted. x represents any amino acid except cysteine, glycine or proline; z represents any of the amino acids asparagine, histidine or tyrosine.



Supplemental Figure S2 - Competition BIAcore of the high-affinity APH inhibitors. A SA-Chip was coated with 5000 RU of biotinylated APH. (**A**) AR_3a at a concentration of 4 nM or (**B**) AR_3b at a concentration of 2.5 nM, preincubated with unbiotinylated APH at different concentrations (0, 1, 2, 3, 4, 5, 6, 7, 8, 12, 16, 20 nM) were run over the chip and on-rates were recorded. After subtraction of a blank reference cell, the on-rates (slope) were determined, plotted against the inhibitor concentration and fitted using Sigma Plot (SPSS Inc; Chicago, IL, USA) as described previously (33). The K_D determined for (**A**) AR_3a was 0.81 ± 0.09 nM and for (**B**) AR_3b 0.54 ± 0.07 nM, respectively.

Structure, Vol. 13, 1131–1141, August, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.04.020

Allosteric Inhibition of Aminoglycoside Phosphotransferase by a Designed Ankyrin Repeat Protein

Andreas Kohl,[1,2] Patrick Amstutz,[1,2] Petra Parizek,[1,2] H. Kaspar Binz,[1] Christophe Briand,[1] Guido Capitani,[1] Patrik Forrer,[1] Andreas Plückthun,[1,*] and Markus G. Grütter[1,*]
[1]Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Switzerland

Summary

Aminoglycoside phosphotransferase (3′)-IIIa (APH) is a bacterial kinase that confers antibiotic resistance to many pathogenic bacteria and shares structural homology with eukaryotic protein kinases. We report here the crystal structure of APH, trapped in an inactive conformation by a tailor-made inhibitory ankyrin repeat (AR) protein, at 2.15 Å resolution. The inhibitor was selected from a combinatorial library of designed AR proteins. The AR protein binds the C-terminal lobe of APH and thereby stabilizes three α helices, which are necessary for substrate binding, in a significantly displaced conformation. BIAcore analysis and kinetic enzyme inhibition experiments are consistent with the proposed allosteric inhibition mechanism. In contrast to most small-molecule kinase inhibitors, the AR proteins are not restricted to active site binding, allowing for higher specificity. Inactive conformations of pharmaceutically relevant enzymes, as can be elucidated with the approach presented here, represent powerful starting points for rational drug design.

Introduction

Bacteria have developed a variety of pathways and mechanisms by which to inactivate antibiotics. Aminoglycosides, including kanamycin, amikacin, and streptomycin, are widely used in hospital care today, and, therefore, bacterial resistance to these antibiotics is a major concern in the health care field (Boehr et al., 2003). In general, aminoglycosides interact with the bacterial ribosome, thereby disrupting protein synthesis of the target cell. One way by which resistance to aminoglycosides can be achieved is by phosphorylation of hydroxy groups of the antibiotic. The aminoglycoside phosphotransferase (3′)-IIIa (APH) is a prototype enzyme for this mechanism (Boehr et al., 2001). It was found to mediate aminoglycoside resistance in *Enterococci* and *Staphylococci*. The crystal structure of APH has been determined (Hon et al., 1997), followed by a very detailed functional and structural analysis of the enzymatic mechanism (Thompson et al., 1999, 2002; Boehr et al., 2001; Burk et al., 2001). The fold of APH is structurally homologous to the catalytic domain of the eukaryotic Ser/Thr/Tyr protein kinases (EPKs; for example, the mitogen-activated protein kinase p38 [Wang et al., 1997], the insulin receptor kinase [Hubbard et al., 1994], or the Abelson tyrosin kinase [ABL] [Schindler et al., 2000]). It was concluded that APH and EPKs share a common ancestor, which suggests that APH is an ideal model system for all kinases sharing this fold (Hon et al., 1997).

EPKs are of great biological and medical importance because of their fundamental role in signal transduction and regulatory pathways in eukaryotic cells. Diseases, including cancer, inflammation and diabetes, are often directly linked to the malfunctioning of EPKs (Noble et al., 2004). The human genome encodes a total of 518 kinases (Manning et al., 2002). The highly conserved catalytic domain of kinases consists of an N-terminal, mostly β sheet-containing lobe and a C-terminal, α-helical lobe. The ATP binding pocket is located in the groove between the two lobes. Most of the known EPK inhibitors bind in the conserved ATP binding site, and a degree of specificity is only achieved by making use of small individual differences in the ATP binding pockets. This often leads to cross-reactivity and unwanted side effects of EPK inhibitors (Noble et al., 2004). Alternative inhibition mechanisms, not directly or exclusively targeting the active site, have proven to be highly efficient, albeit very hard to achieve. A prime example is the drug Gleevec, which only partially binds in the ATP pocket and inhibits ABL by stabilizing an inactive conformation of the kinase (Schindler et al., 2000).

We have developed proteinaceous APH inhibitors based on designed ankyrin repeat (AR) proteins (Amstutz et al., 2005). Designed AR proteins (Binz et al., 2003; Forrer et al., 2003, 2004) are built from single 33 amino acid repeat modules, which stack together to form elongated protein domains (Sedgwick and Smerdon, 1999). Molecules selected from designed AR protein libraries have been shown to bind different target proteins with high affinity and specificity (Binz et al., 2004). Furthermore, AR proteins are highly stable, well-expressed, and do not contain disulfide bonds, allowing for intracellular applications. The selected APH inhibitors (Amstutz et al., 2005) inhibit the enzyme both in vitro and in vivo and confer kanamycin and amikacin sensitivity to a level comparable to the gene knockout.

Here, we describe the crystal structure of APH in complex with one of the most potent AR protein inhibitors, AR_3a, to 2.15 Å resolution. It shows that the AR protein binds to the C-terminal lobe of APH outside the substrate binding pocket and stabilizes a significantly altered APH conformation with a distorted active site. Based on the structural data combined with the kinetic measurements, we suggest an allosteric inhibition mechanism. A comparison to other known allosteric protein kinase inhibitors reveals similarities and differences in the inhibition mechanism.

*Correspondence: gruetter@bioc.unizh.ch (M.G.G.); plueckthun@bioc.unizh.ch (A.P.)
[2]These authors contributed equally to this work.

Structure
1132

Table 1. Statistics of the Data Collection and Refinement of the mAPH/AR_3a Complex

Data Collection	
Space group	$P2_1$
Cell dimensions, Å	a = 59.67, b = 98.08, c = 81.30 $\alpha = \gamma = 90.0$, $\beta = 110.01$
Resolution limits, Å	20.0–2.15
Observed reflections	total: 163,030; unique: 46,440
Completeness, %	97.2 (81.0)[a]
Redundancy	3.5
R_{sym} (% on I)	5.3 (34.7)[a]
Wilson B factor, Å^2	45.0
Refinement	
Resolution range, Å	20.0–2.15
R factor/R_{free}, %	19.9/26.0
Ordered water molecules	353
Rms deviation from ideal geometry	
Bond lengths, Å	0.016
Bond angles, °	1.55
Average B factor, Å^2	52.8

[a] Numbers in parentheses refer to the highest-resolution shell.

Results and Discussion

Structure Determination and Overall Structure of mAPH/AR_3a Inhibitor Complex

To obtain crystals of APH in complex with the selected AR protein inhibitor, AR_3a, a mutant APH (mAPH) was created in which the two surface cysteine residues (C19, C156) of APH were replaced by serine. In wild-type APH (wtAPH), both cysteines can form intermolecular C19-C156 disulfide bonds, which mediate the formation of a homodimer in vitro (Hon et al., 1997). The protein complex mAPH/AR_3a was purified and crystallized in the presence of ATP, and X-ray diffraction data to 2.15 Å resolution were collected as described in Experimental Procedures. The structure was determined by molecular replacement by using wtAPH (Hon et al., 1997) and the designed AR protein E3_5 (Kohl et al., 2003) as search models (see Experimental Procedures). The final model of the mAPH/AR_3a enzyme inhibitor complex has an R factor of 19.9% and an R_{free} of 26.0% (Table 1 and Experimental Procedures).

Even though the protein is inactive in the complex with the AR protein (see below), the overall fold of mAPH is very similar to the one seen in all wtAPH structures (Hon et al., 1997), and it has structural homology to EPKs (see Figure 1). Between the N-terminal lobe, consisting mainly of β sheets, and the C-terminal, mostly α-helical lobe, there is a deep cleft harboring the ATP binding site (Figure 2A). The substrate binding pocket is adjacent to the ATP binding site and is formed mainly by the C-terminal, α-helical lobe. The AR protein inhibitor AR_3a with the typical AR domain fold (Kohl et al., 2003) binds to the α-helical lobe of mAPH. The conformation of this C-terminal lobe differs significantly from the one seen in structures of wtAPH alone. Strikingly, AR_3a binds the C-terminal lobe of mAPH on the opposite side as compared to the substrate (Figure 2A).

mAPH/AR_3a Conformations in the Asymmetric Unit

The asymmetric unit of the mAPH/AR_3a complex crystals contains two full heterodimeric complexes, named AB and CD. A and C represent mAPH molecules, whereas B and D represent AR_3a proteins. In the crystal, mAPH forms a pseudodimer, in which the molecules are related by a non-perfect 2-fold axis (150° rotation; 10° tilt) (Figure 2B) instead of the perfect 2-fold axis in the wtAPH structure (Figure 2C). The pseudodimer contacts cover about 1350 Å^2 buried accessible surface area (ΔASA) per mAPH (Table 2). However, wtAPH and mAPH employ the same set of residues to form the dimer interface (Figures 2B and 2C). The two heterodimeric complexes, AB and CD, exhibit different conformations (Figure 2D). A comparison of the structures of the A and C mAPH molecules with wtAPH reveals that molecule A has a higher similarity to wtAPH than does molecule C. In addition, molecule C has a distorted ATP binding site with the side chains Glu24, Gly25, and Met26 oriented differently. It also shows only weak electron density for the bound ADP at the position of the β phosphate and for the coordinating second Mg^{2+} ion. In contrast, the ATP binding site in molecule A is identical to that of wtAPH. The APH/AR_3a interface of complex AB displays 12 defined H bonds, compared to only 7 in the CD complex (Table 2 and Tables S1 and S2 in the Supplemental Data available with this article online). Therefore, the AB complex is better suited for the following analysis of the structure. The different conformations of APH in the crystal indicate that APH can adopt several different conformations, a phenomenon well known for EPKs (Huse and Kuriyan, 2002; Ozen and Serpersu, 2004).

mAPH/AR_3a Complex Structure

In the following section, we analyze the structure of mAPH, AR_3a, and the mAPH/AR_3a complex in detail. While the global fold of mAPH is conserved, the main differences in the mAPH structure, compared to wtAPH, are located in the AR_3a binding region, including mainly α helices A and B (Figures 1 and 3; see below) and the adjacent aminoglycoside positioning loop. In the wtAPH homodimer, this loop forms part of the dimer interface, which is stabilized in the crystal by two intermolecular disulfide bridges located on β strand one (C19) and on the aminoglycoside positioning loop between α helices A and B (C156). These disulfide bonds probably form only upon isolation of the protein, as this is a cytoplasmatic enzyme. In mAPH, with these cysteines mutated to serine, dimer formation in this region is not observed, and the aminoglycoside positioning loop is much more flexible and probably more similar to the situation seen in solution. In the mAPH-A molecule, this loop could not be modeled due to missing electron density for residues 149–166 (Figure 2A). In the mAPH-C molecule, electron density is visible for this section, but it is not continuous throughout the loop. Interestingly, in the dimeric wtAPH structure, the aminoglycoside positioning loop adopts a short α-helical conformation, whereas in the mAPH-C molecule, a short antiparallel β sheet is formed instead. These results demonstrate the high flexibility for the aminoglycoside positioning loop of monomeric APH in solution.

Superposition of molecules B and D (AR_3a molecules) with the designed AR protein E3_5 (Kohl et al.,

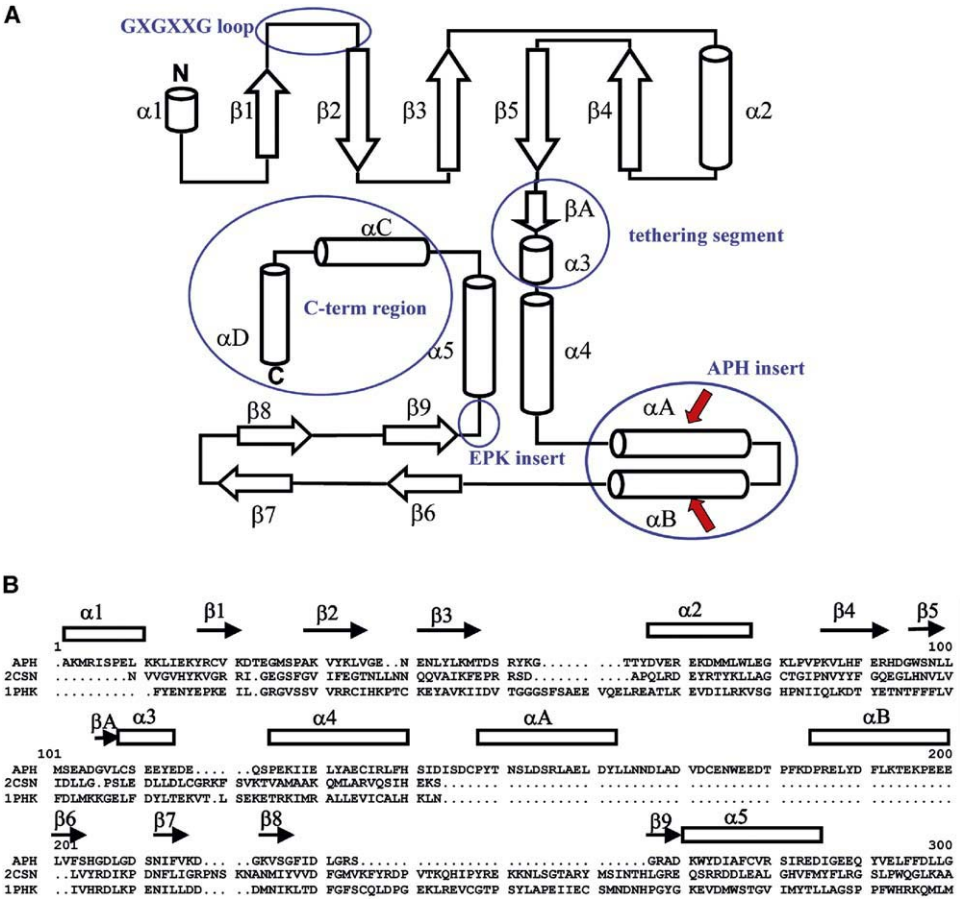


Figure 1. Topology and Structural Alignment of APH and Eukaryotic-Type Protein Kinase

(A) Topological diagram of APH and comparison with Eukaryotic-Type Protein Kinases (EPK). The contact regions between the APH molecule and the ankyrin repeat protein are indicated with red arrows. Blue circles indicate differences in folding between APH and EPK.

(B) Structural alignment between APH and two EPK representatives: the catalytic domain of casein kinase-1 (2CSN) and the catalytic domain of the phosphorylase kinase (1PHK). The C-terminal region of APH after helix α5 bears no resemblance to those seen in EPK's and therefore is not shown in the alignment. (Figure derived from [Hon et al., 1997]).

2003) and the maltose binding protein binder off7 (Binz et al., 2004) reveals virtually identical structures ($rmsd_{C\alpha} \leq 1.5$ Å, Figure 3F). The main differences occur in the C-terminal capping repeat of AR_3a, where weak electron density and high B factors were observed.

The protein-protein interaction surface of the APH/AR_3a complexes was analyzed by the same methods as those previously described (Binz et al., 2004). The analysis revealed a protein-protein interaction interface with global parameters highly comparable to other known AR protein-target protein complexes and other heterodimeric protein-protein interfaces (Lo Conte et al., 1999). The buried accessible surface area (ΔASA) on the AR_3a in the AB complex covers 850 $Å^2$, with 12 H bonds located in the interface (Table 2). Even though there are some differences found in the AB versus CD complex (Table 2), the interacting residues are highly conserved (Tables S1 and S2). The main hydrophobic contacts and six H bonds are formed by the same set of amino acids in both interfaces (Table 2). With 950 $Å^2$ ΔASA, the interaction surface of the AR_3a in the CD complex is even slightly larger than in the AB heterodimer. The differences in the protein-protein interaction result from the different crystal packing for the two heterodimers in the asymmetric unit, as discussed above.

Structural Differences of Free and Inhibited APH

On the mAPH (A and C) surface, mainly residues in α helices A and B are involved in the interaction with the AR protein (Tables S1 and S2). On the AR_3a surface, the interaction residues are located on the N-terminal capping repeat and on repeat modules 1, 2, and 3. While the fold of the AR protein is not affected by the binding (Figure 3F), the structure of mAPH is significantly distorted locally (Figures 3A–3C). α helices A and B in the mAPH model are shifted by 5–7 Å in comparison to the wtAPH structure (Figures 3B and 3C). In addition, α helix B is rotated by 45°–90°, depending on which reference point is chosen (Figure 3C). For example, Tyr172 is partially buried and is only slightly exposed to the solvent in the wtAPH structure, whereas in the mAPH/AR_3a complex, it is involved in the protein-protein interaction with the AR protein and is rotated by

Structure
1134

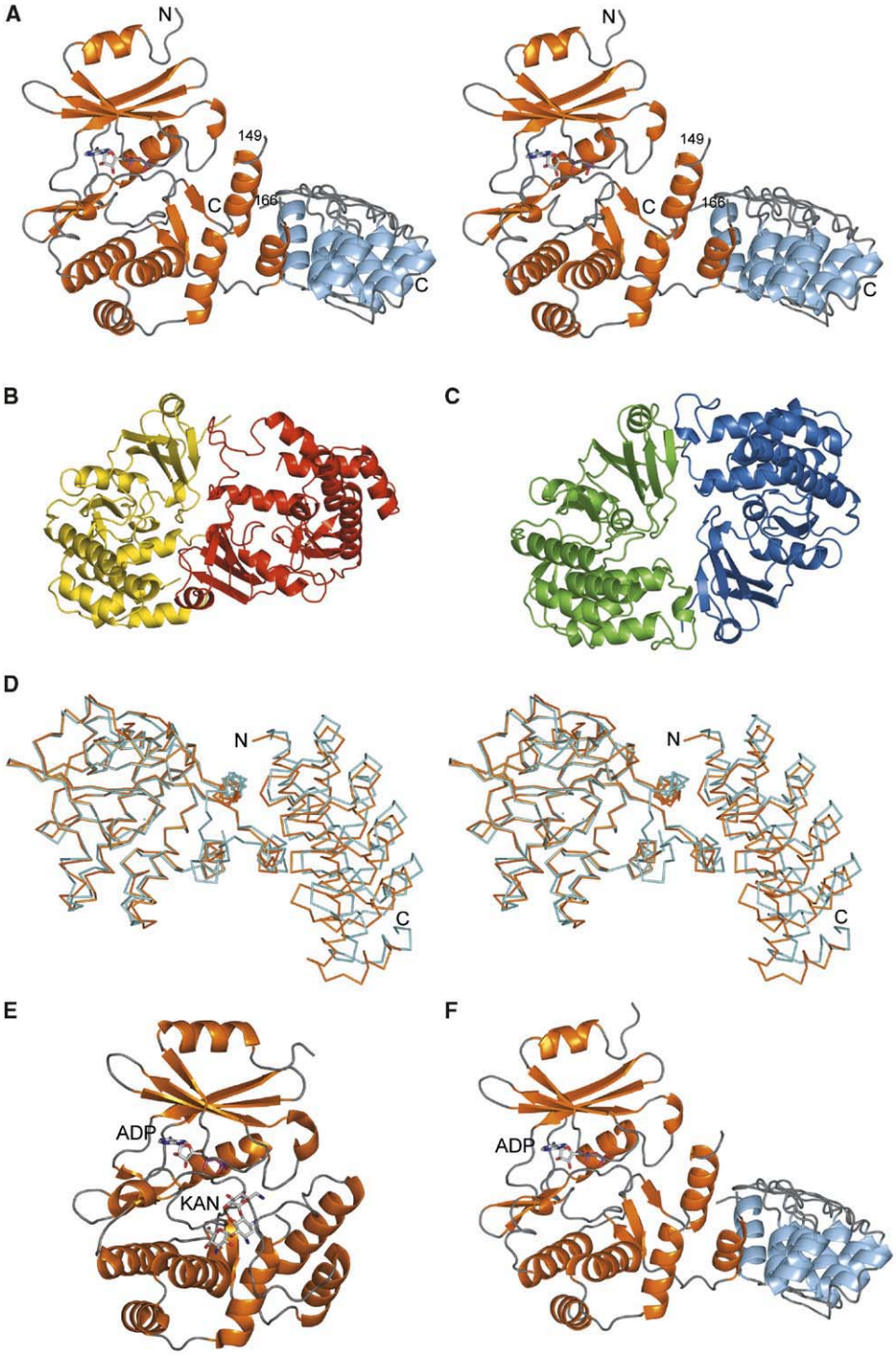


Figure 2. Crystal Structure of the mAPH in Complex with the AR Protein AR_3a

(A) Stereo view of the AB heterodimer of the mAPH/AR_3a complex. The mAPH is shown in orange, and the AR_3a is shown in light blue. AR_3a binds to the C-terminal lobe of the mAPH and stabilizes an inactive conformation.
(B) Asymmetric mAPH pseudo-homodimer found in the crystal of the mAPH/AR_3a complex.
(C) Symmetric wtAPH dimer found in the wtAPH crystal.
(D) Stereo view of the superposition of the mAPH hetereodimer AB (orange) and BC (light blue).
(E) wtAPH in the kanamycin A bound form (PDB: 1L8T). ADP and kanamycin A are labeled.
(F) mAPH/AR_3a AB heterodimer in the same orientation as the wtAPH in (E).

Table 2. Comparison between AR Protein Target Complexes and Other Protein-Protein Interactions

PDB ID[a]	Resolution (Å)	ΔASA[b] (Å^2)	Number of H Bonds	Number of H Bonds/ 100 (Å^2) ΔASA[b,c]	Number of Salt Bridges[c]	Planarity[c]	Number of Bridging H_2O[c]
1awc	2.2	853.62	5	0.58	0	2.30	3
1bi7	3.4	1205.5	7	0.29	4	2.50	0
1blx	1.9	845.4	11	1.30	1	2.50	12
1g3n	2.9	843.3	12	1.42	1	2.20	0
1svx	2.3	611.2	6	1.00	0	2.10	0
APH/AR_3a, AB heterodimer	2.15	847.9	12	2.10	0	2.2	0
APH/AR_3a, CD heterodimer	2.15	950.6	7	0.73	0	2.0	0
APH, AC homodimer	2.15	1371.7	3	0.22	2	3.0	0
APH, CA homodimer	2.15	1367.6	3	0.22	2	3.0	0
Heterodimeric protein-protein complexes[c]	—	983 ± 582	—	1.1 ± 0.5	—	2.8 ± 0.9	—

[a] Protein Data Bank (PDB) accession code (Berman et al., 2000).
[b] Surface area per molecule occluded upon complex formation; if not stated otherwise, the AR protein is analyzed.
[c] According to Lo Conte et al. (1999).

almost 90°. In the wtAPH structure, α helix D is oriented through a number of contacts mainly formed with residues located on α helix B. As a consequence of the alteration of α helix B by AR_3a, α helix D is severely distorted, compared to the wtAPH conformation (see below).

In another recently determined crystal structure of the selected AR protein off7 in complex with maltose binding protein, no structural change of the target protein upon binding of the AR protein was observed (Binz et al., 2004). We therefore believe that APH has a malleable fold. We suggest that AR_3a traps APH in a con-

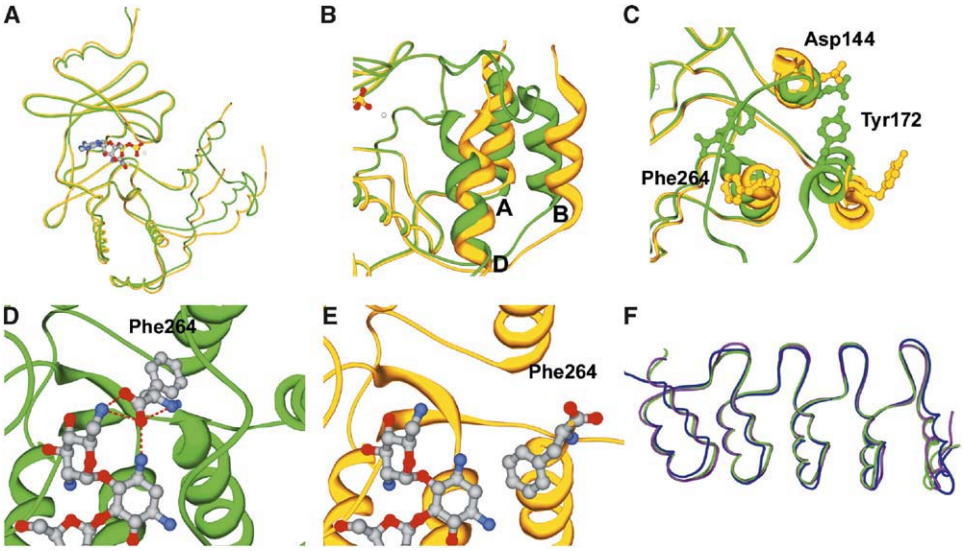


Figure 3. Superpositions and Comparison of wtAPH to mAPH/AR_3a

wtAPH is shown in green, and mAPH/AR_3a (AB heterodimer) is shown in yellow.
(A) Global superposition of the wtAPH structure and the mAPH.
(B) Closeup view of the α helices A, B, and D in the C-terminal lobes of the two APH molecules. The α helices A and B are involved in the interaction with the AR_3a. In mAPH/AR_3a, the α helix D is not held in position anymore by the contacts with α helices A and B. Instead, it is shifted by 5–7 Å with respect to the wtAPH structure.
(C) Closeup view of the α helices A, B, and D rotated by 90° around the horizontal axis with respect to the view in (B). To illustrate the changes in more detail, the side chains of Asp144, Tyr172, and Phe264 are displayed.
(D) Kanamycin A binding pocket in the wtAPH structure and H bonding to Phe264 on α helix D. Phe264 forms three H bonds to kanamycin A.
(E) Disrupted binding site of kanamycin A in the mAPH/AR_3a complex. The kanamycin A is modeled into the mAPH/AR_3a structure. Phe264 on α helix D is not capable of forming H bonds to kanamycin A anymore.
(F) Superposition of the AR proteins E3_5 (green), off7 (purple), and AR_3a (blue). In all three AR proteins, the global fold is unchanged, but local changes occur, especially in the N- and C-terminal capping repeats.

Structure
1136

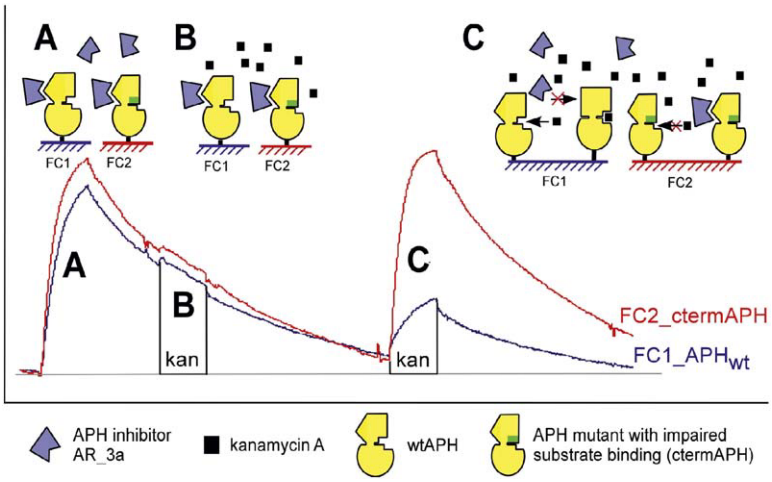


Figure 4. BIAcore Analysis of the Interaction of Inhibitor AR_3a to wtAPH and Inactive ctermAPH in the Presence and Absence of Kanamycin A

Equal amounts of wtAPH in flow-cell one (FC1) (blue) and of ctermAPH in flow-cell two (FC2) (red) were immobilized on a sensor chip. By the additional two C-terminal residues, the active site of ctermAPH is occluded (indicated by a green rectangle in the cartoon).

(A) Binding of the inhibitor was monitored in the absence of kanamycin A, showing comparable binding on both forms of APH.

(B) Addition of a pulse of kanamycin A (200 μM) during the washing procedure did not affect the off-rate of the inhibitor.

(C) Addition of kanamycin A (200 μM) during the binding phase reduced the binding of the inhibitor to the wtAPH, but did not affect the binding to the mutant ctermAPH, which does not bind kanamycin A.

formation that is less populated and perhaps present in solution.

BIAcore Analysis

To investigate whether the binding to kanamycin or to the AR inhibitor is mutually exclusive, the binding behavior of AR_3a to APH was characterized for its dependence on the presence of substrate by BIAcore experiments. The C-terminal amino acid Phe264 of wtAPH, in particular its terminal carboxy group, is essential for substrate binding, as mutations at this site drastically impair binding of substrate (Thompson et al., 1999). We constructed an APH mutant (ctermAPH) with two additional C-terminal amino acids (Gln, Ala). This mutant showed no kinase activity.

Similar amounts of both wtAPH and ctermAPH were immobilized on parallel flow cells on a BIAcore chip. Binding of AR_3a to wtAPH and ctermAPH was monitored in parallel. A pulse of 200 μM kanamycin A was applied either during the injection phase, showing the effect of kanamycin A on the association rate, or during washing, showing the kanamycin A effect on the dissociation rate (Figure 4). As expected, the association and dissociation of AR_3a to ctermAPH—which does not interact with kanamycin A—is identical in the presence or absence of kanamycin A. The association of AR_3a to wtAPH, however, is strongly reduced in the presence of kanamycin A (Figure 4). The apparent on-rate was reduced more than 150-fold in the presence of 100 μM kanamycin A (data not shown). The off-rate, on the other hand, was not influenced by kanamycin A (Figure 4). In conclusion, kanamycin A and AR_3a seem to compete for binding to wtAPH, albeit not for the active site, but rather for a malleable APH molecule that can assume different conformations. The dissociation of the APH/AR_3a complex occurs with a given rate, which seems not to be influenced by kanamycin A.

Inhibition

To further elucidate the mechanism of inhibition of the inhibitor AR_3a, in vitro inhibition studies were performed. wtAPH, AR_3a, and wtAPH/AR_3a complex were expressed and purified as described in Experimental Procedures. wtAPH purification resulted in a mixture of monomer and dimer as described previously (McKay et al., 1994). For the in vitro inhibition studies, only the isolated monomeric APH was used. Purified AR_3a protein tended to form a mixture of monomers, dimers, and multimers, which showed a fast equilibrium (Amstutz et al., 2005). This is a property of this particular inhibitor (Binz et al., 2003, 2004). The isolated wtAPH/AR_3a complex was investigated by size-exclusion chromatography and gave rise to a single peak with a 1:1 ratio (Amstutz et al., 2005).

The steady-state velocities of enzyme activity were monitored by coupling the release of ADP to a pyruvate kinase/lactate dehydrogenase reaction as described previously (McKay et al., 1994), with some modifications (see Experimental Procedures). We decided to use amikacin for these studies, as, with this substrate, APH follows simple Michaelis-Menten kinetics. This is in contrast to what is seen in kanamycin A, a case in which the analysis is complicated by substrate inhibition (Amstutz et al., 2005). Due to the nature of the coupled assay and the lack of knowledge of the exact molarity of the active species of both wtAPH and AR_3a, a quantitative analysis of the inhibition mechanism was not possible. Therefore, we focused on a qualitative investigation of the inhibition mechanism and give ranges of constants when possible.

For determining the binding mode, progress curves in the presence of wtAPH alone and wtAPH at various concentrations of the inhibitor AR_3a were analyzed. If the reaction was started with the enzyme, each curve with inhibitor present showed an initial exponential "burst" phase, followed by a slower steady-state rate after some minutes (data not shown). These curves suggest slow, tight binding inhibition. The association and dissociation rate constants (k_{on} = 1.6•10^6 M^{-1} s^{-1}, k_{off} = 2.7•10^{-3} s^{-1}) determined in BIAcore experiments (Amstutz et al., 2005) further support a tight binding mechanism with observable pre-steady-state kinetics in this range.

For the classification of the inhibition mechanism,

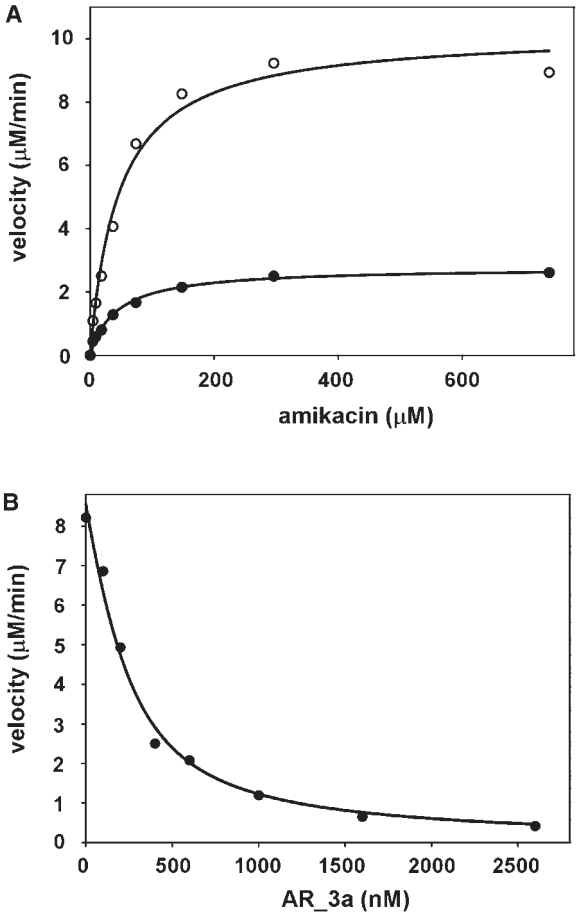


Figure 5. In Vitro Inhibition of wtAPH by AR_3a

(A) Enzyme activities of both wtAPH (open circles) and purified wtAPH/AR_3a complex (closed circles) were analyzed at various amikacin concentrations. The wtAPH/AR_3a complex shows a reduction of V_{max} with little effect on the K_m value.

(B) Amikacin phosphorylation by wtAPH was determined in the presence of various concentrations of the inhibitor AR_3a. Steady-state velocities were measured at 740 μM amikacin and were plotted as a function of inhibitor concentration. wtAPH seems to be completely inhibited by AR_3a. All assays were carried out at 20°C with 200 nM wtAPH or 200 nM equilibrated wtAPH/AR_3a complex.

steady-state velocities of wtAPH and wtAPH/AR_3a complex were determined at various substrate concentrations, and the steady-state velocities were plotted as a function of substrate concentration (Figure 5A). The rates were fitted by using the standard Michaelis-Menten equation. The K_m and V_{max} values determined for 200 nM wtAPH are 46.43 ± 7.16 μM and 10.22 ± 0.45 μM/min (corresponding to k_{cat} = 0.85 s^{-1}), respectively, and are 42.41 ± 4.63 μM and 2.78 ± 0.08 μM/min (corresponding to k_{cat} = 0.23 s^{-1}), respectively, for 200 nM equilibrated wtAPH/AR_3a complex. The wtAPH/AR_3a complex thus shows a reduction of V_{max} with little effect on the K_m value. This is in agreement with a mixed-type inhibition (Scheme 1), containing a competitive and an uncompetitive component. A purely uncompetitive inhibition, which would predict a lower apparent K_m, can furthermore be ruled out, because AR_3a also binds the free enzyme in BIAcore experiments (Figure 4). A purely competitive inhibition would not alter V_{max}. In Figure 5B, velocities of both wtAPH alone and in the presence of various concentrations of AR_3a were plotted as a function of inhibitor concentration. Inhibitor titration revealed that AR_3a completely inhibits the enzymatic activity of wtAPH in vitro. This is in agreement with the structural data, by which the substrate binding site is shown to be highly disordered upon inhibitor binding (see below). Virtually identical results were obtained with mAPH (data not shown), suggesting that the introduction of the two cysteine to serine mutations had no influence on the enzyme or on its inhibition by AR_3a.

For the determination of the inhibition constant αK_i, the data were fitted to the equation for tight binding conditions (Scheme 1; Szedlacsek et al., 1988) or directly with the model of Scheme 1 by DYNAFIT (Kuzmic, 1996). In either case, the αK_i value obtained was 140 ± 18 nM, with the α coefficient in the range of 1–3. Since the active species of enzyme and inhibitor could not be determined, a separate specification of K_i and α was not possible. Yet estimating α = 1–3 from the fit, the K_i obtained in the in vitro inhibition studies is 30–100 times higher than the K_D determined in BIAcore experiments (K_D = 1.7 nM; Amstutz et al., 2005). There are several contributing factors to this observation. First, in the BIAcore measurements, only the interaction between APH and AR_3a is investigated, whereas in the kinetic studies, the substrates amikacin and Mg•ATP are also present. The second reason might result from different assay conditions (different buffer composition) of the kinetic measurements. Third, overestimation of the concentration of active AR_3a molecules due to oligomerization (Amstutz et al., 2005) in the kinase buffer might have an effect on the observed inhibition constant.

Even though we cannot quantitatively determine the kinetic parameters, it is important to emphasize that these in vitro inhibition studies demonstrate that AR_3a completely inhibits APH (wtAPH and mAPH) by a mixed-type inhibition mechanism.

Inhibition Mechanism

Based on the structure of the complex and the in vitro data on inhibition and binding, we propose the following inhibition mechanism: the AR_3a inhibitor and the substrate compete for binding to APH. The conformation stabilized by the inhibitor is unproductive, while the one binding substrate is the active form of APH. The mixed-type inhibition seen kinetically (Figure 5A) suggests that the substrate can bind, albeit poorly, to the enzyme-inhibitor complex, EI; however, the complete inhibition at high inhibitor concentrations (Figure 5B) shows that the enzyme-substrate-inhibitor complex (ESI) is not catalytically active.

The reaction mechanism for the phosphorylation of kanamycin by wtAPH was analyzed in detail (Hon et al., 1997; Thompson et al., 1999, 2002; Boehr et al., 2001; Burk et al., 2001). It was shown that, among other residues, the very C-terminal amino acid Phe264 in α helix D is crucial for substrate binding (Figure 3D) (Thompson et al., 1999). The terminal carboxylate of Phe264 forms

Structure
1138

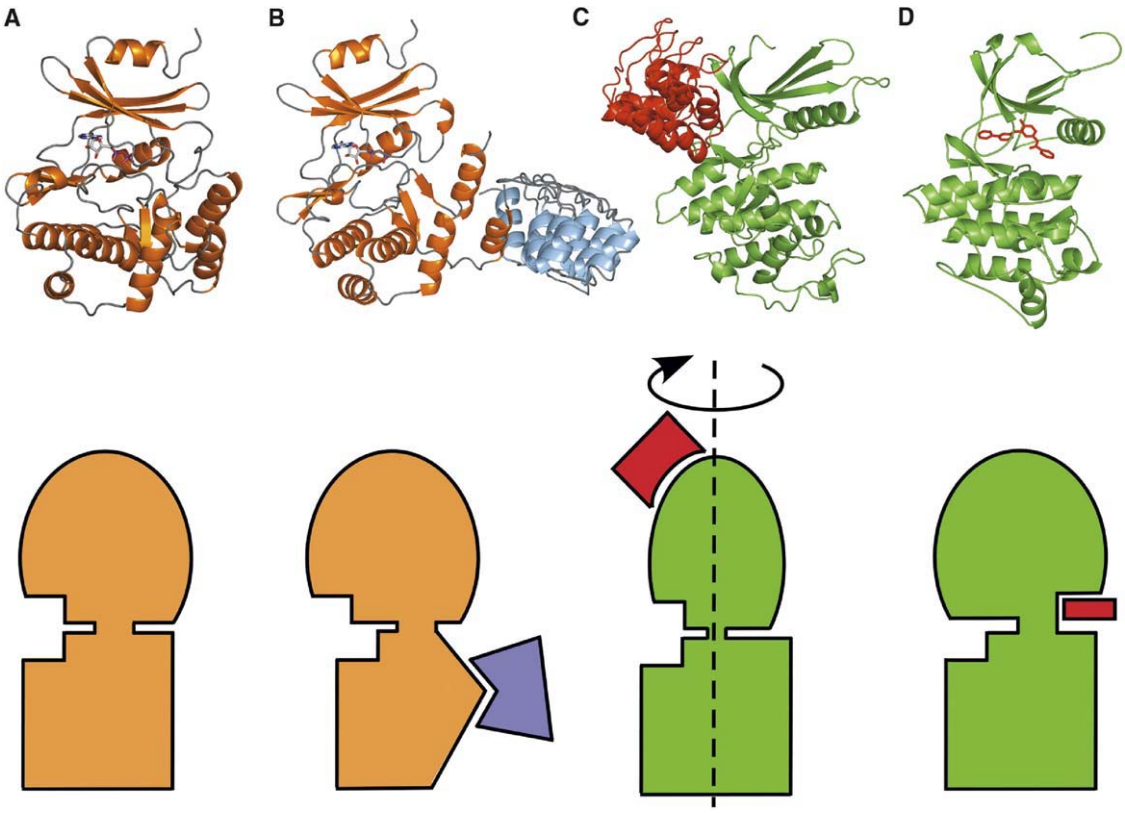


Figure 6. Structural Comparison of mAPH/AR_3a Structure to Other Inhibitor-Kinase Complexes

(A) The wtAPH monomer is shown in orange.
(B) mAPH/AR_3a structure (AB heterodimer) with mAPH, shown in orange, bound by AR_3a, shown in light blue. The AR protein binds to the C-terminal lobe of the kinase and distorts the kanamycin A binding site.
(C) Cyclin-dependent kinase 6 shown in green, bound by AR protein p18, shown in red. p18 binds to the N-terminal lobe and distorts residues in the ATP binding pocket, altering the binding surface for the cyclin substrate.
(D) Crystal structure of the Abelson tyrosine kinase, shown in green, and the inhibitor STI-571 (Gleevec), shown in red. STI-571 stabilizes an inactive conformation of ABL, in which the central activation loop remains unphosphorylated; therefore, the kinase is inactive.

three H bonds to the substrate and thus orients the aminoglycoside in the active site. Point mutations or deletions of Phe264 turned out to have a drastic effect on the enzymatic activity, impairing the binding of kanamycin to APH and thus inhibiting the enzyme (Thompson et al., 1999). In the mAPH/AR_3a complex, we observe that, by binding of the AR protein to mAPH, α helices A and B are oriented differently than in the wtAPH structure. In the wtAPH structure, the α helices A and B and the aminoglycoside positioning loop are needed to form a shell that positions the C-terminal α helix D. In the inhibited mAPH/AR_3a complex, this shell is removed and α helix D is much more flexible. This leads to a shift of up to 7 Å and a movement of the C-terminal α helix D (Figures 3B and 3E). As a consequence, the terminal carboxylate of Phe264 cannot form the essential H bonds with kanamycin anymore, finally resulting in enzyme inhibition (Figure 3E). The inhibition data are in agreement with the effects seen for Phe264 point mutations and deletions (Thompson et al., 1999) and also confirm the inactivity of the mutant extended by two amino acids (data not shown). BIAcore experiments demonstrate that kanamycin A, which interacts with the terminal carboxylate of Phe264, stabilizes a conformation, which is only poorly recognized by AR_3a. All results indicate that inhibition of APH is achieved through stabilization of an inactive conformation of APH by AR_3a.

Comparison to Other AR Protein Kinase Inhibitors and Gleevec: Implications for Drug Design and Cell Biology

Natural AR proteins are well known as regulators and inhibitors of kinases. Several crystal structures of the cyclin-dependent kinase 6 (Cdk6) with different AR protein inhibitors of the INK4 family (p16, p18, and p19) illustrate an allosteric inhibition mechanism (Russo et al., 1998; Brotherton et al., 1998; Pavletich, 1999; Jeffrey et al., 2000). Upon binding of the AR protein, a conformational change occurs in Cdk6. The N-terminal lobe of Cdk6 rotates by 10°–15°, leading to a misalignment of residues needed to bind ATP and to a conformational change in the interface needed for the binding of cyclin (Figure 6C), which in turn is needed to activate the kinase. The AR proteins p16, p18, and p19 bind mainly to the N-terminal lobe of Cdk6, at the back of the ATP binding cleft. The inhibition mechanism seen in the mAPH/AR_3a complex is remarkably similar, even though the AR protein binds to a different part of the protein. The binding of AR_3a to the C-terminal lobe of

APH induces a structural change in the substrate binding site, leaving the ATP binding pocket unchanged, but resulting in an inhibited APH (Figures 6A and 6B).

As shown here and in Amstutz et al. (2005), selected AR protein inhibitors for kinases are highly specific and do not need to target the ATP binding site for inhibition, but exploit "natural" conformational fluctuations of kinases. Most conventional small-molecule kinase inhibitors target the ATP binding pocket and act as competitive inhibitors to ATP. As this site is highly conserved, specificity is difficult to achieve (Noble et al., 2004). For example, inhibition of APH by some EPK inhibitors has been described (Daigle et al., 1997). This emphasizes the structural homology of APH and EPKs, especially for the ATP binding site. It also demonstrates the poor specificity of these EPK inhibitors. For human EPKs involved in diseases, specificity is absolutely required. Unspecific binding of inhibitors to several EPKs can cause severe side effects and can prevent their pharmaceutical use (Noble et al., 2004).

The Abelson tyrosine kinase (ABL) inhibitor ST1-571 (Gleevec) binds to the N-terminal lobe of ABL and in the ATP binding pocket, and thereby traps the kinase in an inactive conformation, with its activation loop bound back into the peptide binding cleft (Figure 6D; Schindler et al., 2000). This prevents the phosphorylation of the binding loop and thus the activation of ABL kinase. So far, targeting such an inactive form of a kinase was a pure trial-and-error process and relied on high-throughput screening (Noble et al., 2004). The structure presented here, and especially the approach via the selection of specific protein-based inhibitors, could open new opportunities in the structure-based drug design (Anderson, 2003). The mAPH/AR_3a crystal structure showing a nonproductive kinase conformation, and similar studies for other EPKs, could serve as a starting point for structure-based drug design of APH and other kinase inhibitors. In principle, the same approach of using selected AR protein inhibitors for cocrystallization could also be applied to EPKs. In fact, pools of binders to JNK2 and p38 have already been selected, and the binding of some individual binders has been characterized (Binz et al., 2004). The pools of binders could act as a starting point for inhibitor screens, followed by cocrystallization of identified inhibitor/enzyme complexes. Subsequent structure-based drug design might allow for the development of highly specific drugs. This approach is not at all limited to kinases, and it might be applied to any other enzyme.

In a cell, kinases are not only regulated by direct activation or inactivation, but they are regulated by their localization as well (their interplay with other kinases, their corresponding phosphatases, and other regulating proteins [Sharrocks et al., 2000]; [Enslen and Davis, 2001]; [Pawson and Nash, 2003]; [Tanoue and Nishida, 2003]). Amstutz et al. (2005) showed that AR proteins can be used intracellularly as well as extracellularly, providing a similar phenotype as a knockout mutant. Therefore, AR proteins may also interfere in signaling cascades of kinases or in other pathways. It can be easily envisioned that the binding of a selected AR protein can prevent the docking of the kinase to the scaffolding protein, the interaction with a phosphatase, or the interaction with an upstream activating kinase. The possibility of preventing specific interactions with a selected AR protein would offer individual researchers many novel possibilities to design their experiments and has several advantages over conventional methods used so far (for a more detailed discussion, see [Amstutz et al., 2005]).

Experimental Procedures

Cloning and Production of Proteins

Biotinylated wtAPH was produced as described in Amstutz et al., (2005). A biotinylated mutant of APH (ctermAPH), carrying two additional amino acids at the C terminus (Gln, Ala), was constructed and produced as described for the biotinylated wtAPH (Amstutz et al., 2005), with the sole difference that the reverse primer for the PCR of the APH gene was APHmutCTr (5′-TACTCAAGCTTGAAA CAATTCATCCA-3′). For the construction of the untagged double cysteine to serine mutant (mAPH), in which cysteines 19 and 156 of APH were mutated to serines, the mutations were introduced by PCR. The mAPH gene was inserted into pQE60 (Qiagen, Hilden, Germany) without a tag and expressed in *E. coli* XL-1Blue, as described for the AR proteins (Binz et al., 2003, Kohl et al., 2003). APH variants (wtAPH, mAPH, ctermAPH) were purified by using a 25 ml Q-Sepharose column (Amersham Pharmacia, Dübendorf, Switzerland), followed by preparative Superdex-75 (Amersham Pharmacia) size-exclusion chromatography in 10 mM HEPES (pH 7.5), 100 mM NaCl, 50 mM $MgCl_2$, and 2 mM dithiothreitol (DTT), according to the published protocol (Hon et al., 1997). Only fractions corresponding to the monomeric APH were used for subsequent experiments. For BIAcore and inhibition experiments, the APH inhibitor AR_3a was expressed and purified as described (Binz et al., 2003, Kohl et al., 2003).

mAPH/AR_3a Complex

wtAPH, mAPH, and the selected APH inhibitor AR_3a were expressed as described above. The cell pellets of 1 l bacterial culture of APH and AR_3a were resuspended and combined in 30 ml buffer (50 mM HEPES [pH 7.5], 200 mM NaCl, 10 mM $MgCl_2$). The cells were lysed by using an Emulsiflex C5 (Avestin, Ottawa, ON, Canada) with the addition of Complete Protease Inhibitor (Roche, Basel, Switzerland) and DNaseI (Roche). The cell debris was pelleted by centrifugation at 50,000 × g for 30 min at 4°C. The resulting supernatant was applied to an immobilized metal-ion affinity chromatography column (Qiagen), equilibrated with 50 mM HEPES (pH 7.5), 200 mM NaCl, 20 mM imidazole, 10 mM $MgCl_2$. After extensive washing with equilibration buffer, the complex was eluted with 50 mM HEPES (pH 7.5), 200 mM NaCl, 250 mM imidazole, 10 mM $MgCl_2$ and was further purified by preparative Superdex-75 (Amersham Pharmacia) size-exclusion chromatography in 10 mM HEPES (pH 7.5), 100 mM NaCl, 50 mM $MgCl_2$, and 3 mM DTT. The peak fraction with the expected molecular weight of the complex, containing one of the APH variants and AR_3a in equimolar amounts (determined by SDS-PAGE), was collected and used for further experiments. Dynamic light scattering of these fractions was measured as described (Kohl et al., 2003) and showed a monodisperse distribution for both complexes.

Crystallization

The mAPH/AR_3a complex was concentrated to 10–12 mg/ml, and 5 mM ATP (pH 7.0) was added. Initial crystallization screening was done by using a four channel Lissy (Zinsser, Germany) liquid handler for pipetting the reservoir and a Cartesian eight channel MicroSys (Zinsser) system for the sitting drop setup. As a default, 100 μl reservoir solution was used in combination with 100 nl protein solution, added to 100 nl reservoir solution in 96-well sitting drop CrystalQuick round well crystallization plates (Greiner Bio-One, Frickenhausen, Germany). The initially obtained crystals were refined by using standard techniques in a 1 μl + 1 μl setup. The crystals used for data collection were grown in 1–3 days in 15%–20% PEG 5500, 0.1 M Mes/HCl (pH 5.9–6.2). For cryoprotection, 10% ethylene glycol was employed.

Data Collection, Reduction, Structure Solution, and Refinement

Data of single crystals were collected at the Swiss Light Source biocrystallography beamline (Villigen, Switzerland). The data were processed with DENZO, SCALEPACK, and TRUNCATE (Otwinowski and Minor, 1997). The crystal belonged to space group $P2_1$, with a Matthews coefficient V_M of 2.1 $Å^3$/Da, corresponding to an estimated water content of 41%.

The crystal structure was determined by molecular replacement by using the program AMoRe (Navaza, 1994, 2001), with the structure of the unselected N3C library member E3_5 (PDB-ID 1MJ0 [Kohl et al., 2003]) and a monomeric APH model (PDB-ID 1J7L) used as search models. First, we applied a conventional AMoRe protocol with a wtAPH monomer as a search model. This yielded two solutions for mAPH with only very little additional density for a potential AR protein and no success in further refinement. An AMoRe search with the AR protein E3_5 as a search model did not yield a meaningful solution. The program AMoRe offers the possibility to apply an n-body translational search, which is routinely used in AMoRe, but usually only with one search model. Here, we applied a modified protocol for an n-body search with two search models. Two separate structure factor tables, followed by separate rotational searches for each of the two search models, were calculated from APH (1J7L) and from the AR protein (1MJ0). A conventional translational search was applied for a first APH molecule, which led to a clear solution. This solution was fixed, and the translational search was repeated for a second APH molecule, leading to a second solution, which was again fixed. In the next step, the structure factor table and the rotational search list of the AR protein E3_5 were used in a translational search in combination with the already fixed solutions for the APH, now yielding a solution for the AR protein. A second search with two fixed APH molecules and one fixed AR protein failed to give a solution distinguishable from the background for a second AR protein. However, in the electron density derived, using the phases of two APH and one AR protein, clear additional density was seen for a second AR protein. The second AR protein was modeled into the density with the program O (Jones et al., 1991). Subsequent refinement with CNS (Brünger et al., 1998), REFMAC (Murshudov et al., 1999), the CCP4 program suite (CCP4, 1994), and, for the model building, O (Jones et al., 1991) resulted in a final model with an R factor of 19.9% and an R_{free} of 26.0%, Table 1.

Analysis of the Complexes

The structural analysis of the complexes was done as suggested by Jones and Thornton (1996) and as described previously (Binz et al., 2004).

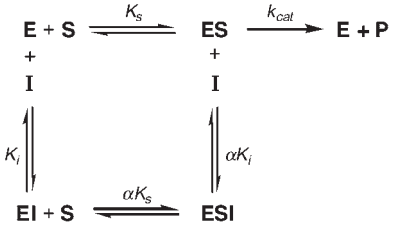
Surface Plasmon Resonance

SPR was measured with a BIAcore3000 instrument (BIAcore, Uppsala, Sweden). The running buffer was 20 mM HEPES (pH 7.4), 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 0.005% Tween 20. An SA chip (BIAcore) was used with about 500 RU biotinylated pD_APH immobilized on flow cell 1 and about 450 RU biotinylated pD_ctermAPH on flow cell 2. AR_3a was applied at a concentration of 200 nM either in running buffer or in running buffer with 200 μM kanamycin A. After application of AR_3a in the absence of kanamycin A, a pulse of 150 μl running buffer containing 200 μM kanamycin A was injected during dissociation (for details, see Figure 4). On- and off-rates in the presence of 100 μM kanamycin A were measured at a flow of 50 μl/min with 5 min buffer flow, a 2 min injection of AR_3a in varying concentrations (10, 50, 100, 200, 500, 1000 nM), and an off-rate measurement of 40 min with buffer flow. The signal of an uncoated reference cell was subtracted from the measurements. The kinetic data of the interaction were evaluated by using the software BIAevaluation 3.0 (BIAcore).

Enzyme Assay

APH activity was monitored by coupling the production of ADP to the consumption of NADH in a pyruvate kinase/lactate dehydrogenase (PK/LDH) reaction as described (McKay et al., 1994), with some modifications (Amstutz et al., 2005). 990 μl assay buffer (50 mM HEPES [pH 7.5], 10 mM $MgCl_2$, 40 mM KCl, 160 mM NaCl, 5U PK, 5U LDH, 2.5 mM phosphoenolpyruvate, 0.025% BSA, 0.12 mg/ml NADH, 1 mM ATP) containing either 200 nM purified wtAPH or wtAPH/AR_3a complex (1:1) was added in a 10 mm QS cell (HELLMA, Basel, Switzerland). The mixtures were preincubated for 15 min, and the assay was initiated by addition of 10 μl of different amikacin concentrations (Sigma, A1774). Both the preincubation and the assay were carried out at 20°C. The oxidation of NADH was followed by continuously monitoring the absorbance at 340 nm. The steady-state rates were normalized for background activity by subtracting the activity monitored in the absence of amikacin. Velocities were plotted as a function of the substrate concentration. The data were fitted to the standard Michaelis-Menten equation to determine V_{max} and K_m values.

For the analysis of the inhibition constant, the kinetic measurements were carried out with 200 nM wtAPH or wtAPH in complex with various amounts of inhibitor AR_3a. For the formation of complexes, 35 μM wtAPH was mixed with AR_3a in a molar ratio of 1:0.5 to 1:13 and was incubated for 2 hr at 4°C prior to the dilution (1:175) in the enzyme assay. The reaction was started with 740 μM amikacin. Enzyme activities were plotted as a function of the inhibitor concentration, and the data were fitted by the general equation for tight binding inhibition based on the mechanism in Scheme 1 (Szedlacsek et al., 1988).



Scheme 1.

In this scheme, E is the enzyme, S represents the substrate, P represents the product, and I represents the inhibitor. K_S = [E][S]/[ES], K_i = [E][I]/[EI], αK_S = [EI][S]/[ESI], αK_i = [ES][I]/[ESI], k_{cat} = catalytic constant, and α is a coefficient. The αK_i value was determined to 140 ± 18 nM with the α coefficient in the range of 1–3.

Supplemental Data

Supplemental Data including the accessible surface area per residue buried and the hydrogen bonds formed in the APH/AR_3a complex are available at http://www.structure.org/cgi/content/full/13/8/1131/DC1/.

Acknowledgments

We would like to thank Sandra Lepthien, Maya Gulotti, and Beat Blattmann for help in protein preparation and crystallization and the members of the Plückthun and Grütter laboratories for valuable discussions. We would also like to thank Prof. Antonio Baici for his help and discussion on APH kinetics. H.K.B. was supported by a predoctoral fellowship of the Roche Research Foundation. This work was supported by the Swiss National Center of Competence in Research (NCCR) in Structural Biology and the Swiss Krebsliga.

Received: February 17, 2005
Revised: April 25, 2005
Accepted: April 26, 2005
Published: August 9, 2005

References

Amstutz, P., Binz, H.K., Parizek, P., Stumpp, M.T., Kohl, A., Grütter, M.G., Forrer, P., and Plückthun, A/ (2005). Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. J. Biol. Chem. *280*, 24715–24722.

Anderson, A. (2003). The process of structure-based drug design. Chem. Biol. *10*, 787–797.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N.,

Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J. Mol. Biol. *332*, 489–503.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grütter, M.G., and Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. Nat. Biotechnol. *22*, 575–582.

Boehr, D.D., Thompson, P.R., and Wright, G.D. (2001). Molecular mechanism of aminoglycoside antibiotic kinase APH(3′)-IIIa. Roles of conserved active site residues. J. Biol. Chem. *276*, 23929–23936.

Boehr, D.D., Draker, K., Koteva, K., Bains, M., Hancock, R.E., and Wright, G.D. (2003). Broad-spectrum peptide inhibitors of aminoglycoside antibiotic resistance enzymes. Chem. Biol. *10*, 189–196.

Brotherton, D., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P., Volyanik, E., Xu, X., Parisini, E., Smith, B., Archer, S., et al. (1998). Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor $p19^{INK4d}$. Nature *395*, 244–250.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Burk, D., Hon, W., Leung, A., and Berghuis, A. (2001). Structural analyses of nucleotide binding to an aminoglycoside phosphotransferase. Biochemistry *40*, 8756–8764.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Daigle, D.M., McKay, G.A., and Wright, G.D. (1997). Inhibition of aminoglycoside antibiotic resistance enzymes by protein kinase inhibitors. J. Biol. Chem. *272*, 24755–24758.

Enslen, H., and Davis, R.J. (2001). Regulation of MAP kinases by docking domains. Biol. Cell. *93*, 5–14.

Forrer, P., Stumpp, M.T., Binz, H.K., and Plückthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett. *539*, 2–6.

Forrer, P., Binz, H.K., Stumpp, M.T., and Plückthun, A. (2004). Consensus design of repeat proteins. ChemBioChem *5*, 183–189.

Hon, W., McKay, G., Thompson, P., Sweet, R., Yang, D., Wright, G., and Berghuis, A. (1997). Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. Cell *89*, 887–895.

Hubbard, S.R., Wei, L., Ellis, L., and Hendrickson, W.A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature *372*, 22–29.

Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell *109*, 275–282.

Jeffrey, P.D., Tong, L., and Pavletich, N.P. (2000). Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. Genes Dev. *14*, 3115–3125.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. D Biol. Crystallogr. *47*, 110–119.

Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Plückthun, A., and Grütter, M.G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc. Natl. Acad. Sci. USA *100*, 1700–1705.

Kuzmic, P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal. Biochem. *237*, 260–273.

Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177–2198.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912–1934.

McKay, G.A., Thompson, P.R., and Wright, G.D. (1994). Broad spectrum aminoglycoside phosphotransferase type III from *Enterococcus*: overexpression, purification, and substrate specificity. Biochemistry *33*, 6936–6944.

Murshudov, G.N., Lebedev, A., Vagin, A.A., Wilson, K.S., and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. D Biol. Crystallogr. *55*, 247–255.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

Navaza, J. (2001). Implementation of molecular replacement in AMoRe. Acta Crystallogr. D Biol. Crystallogr. *57*, 1367–1372.

Noble, M.E.M., Endicott, J.A., and Johnson, L.N. (2004). Protein kinase inhibitors: insights into drug design from structure. Science *303*, 1800–1805.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, C.W. Carter and J. R.M. Sweets, eds. (New York: Academic Press), pp. 307–326.

Ozen, C., and Serpersu, E.H. (2004). Thermodynamics of aminoglycoside binding to aminoglycoside- 3′-phosphotransferase IIIa studied by isothermal titration calorimetry. Biochemistry *43*, 14667–14675.

Pavletich, N.P. (1999). Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. J. Mol. Biol. *287*, 821–828.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445–452.

Russo, A., Tong, L., Lee, J., Jeffrey, P., and Pavletich, N. (1998). Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor $p16^{INK4a}$. Nature *395*, 237–243.

Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., and Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science *289*, 1938–1942.

Sedgwick, S.G., and Smerdon, S.J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem. Sci. *24*, 311–316.

Sharrocks, A.D., Yang, S.-H., and Galanis, A. (2000). Docking domains and substrate-specificity determination for MAP kinases. Trends Biochem. Sci. *25*, 448–453.

Szedlacsek, S.E., Ostafe, V., Serban, M., and Vlad, M.O. (1988). A re-evaluation of the kinetic equations for hyperbolic tight-binding inhibition. Biochem. J. *254*, 311–312.

Tanoue, T., and Nishida, E. (2003). Molecular recognitions in the MAP kinase cascades. Cell. Signal. *15*, 455–462.

Thompson, P.R., Schwartzenhauer, J., Hughes, D.W., Berghuis, A.M., and Wright, G.D. (1999). The COOH terminus of aminoglycoside phosphotransferase (3′)-IIIa is critical for antibiotic recognition and resistance. J. Biol. Chem. *274*, 30697–30706.

Thompson, P.R., Boehr, D.D., Berghuis, A.M., and Wright, G.D. (2002). Mechanism of aminoglycoside antibiotic kinase APH(3′)-IIIa: role of the nucleotide positioning loop. Biochemistry *41*, 7001–7007.

Wang, Z., Harkins, P.C., Ulevitch, R.J., Han, J., Cobb, M.H., and Goldsmith, E.J. (1997). The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution. Proc. Natl. Acad. Sci. USA *94*, 2327–2332.

Accession Numbers

The coordinates and structure factors for the mAPH/AR_3a complex have been deposited in the Protein Data Bank (accession code 2BKK).

Supplemental Data

Allosteric Inhibition of Aminoglycoside

Phosphotransferase by a Designed

Ankyrin Repeat Protein

Andreas Kohl, Patrick Amstutz, Petra Parizek, H. Kaspar Binz, Christophe Briand, Guido Capitani, Patrik Forrer, Andreas Plückthun, and Markus G. Grütter

Table S1. mAPH Residues on Molecules A and C Involved in the Interaction with AR_3a Molecules B and D, Respectively

Residue#	Residue	Chain	Interface ΔASA[a]	% Interface ΔASA[a]	HB[b]	Chain	Interface ΔASA[a]	% Interface ΔASA[a]	HB[b]
136	LEU	A	23.25	2.51		C	17.44	1.82	
137	ASP	A	75.26	8.13	2	C	54.15	5.64	
140	LEU	A	46.03	4.97		C	50.61	5.27	
141	ALA	A	23.22	2.51		C	39.38	4.10	
143	LEU	A	15.06	1.63		C	7.00	0.73	
144	ASP	A	80.87	8.73	2	C	61.20	6.37	1
146	LEU	-	-	-		C	17.46	1.82	
147	LEU	A	13.88	1.50		C	28.16	2.93	
148	ASN	A	1.12	0.12		-	-	-	
164	PRO	-	-	-		C	9.34	0.97	
165	PHE	-	-	-		C	75.94	7.90	1
166	LYS	A	36.98	3.99	1	C	4.66	0.48	
168	PRO	A	51.21	5.53		C	72.11	7.51	
169	ARG	A	108.25	11.69		C	122.56	12.76	
171	LEU	A	44.93	4.85		C	25.03	2.61	
172	TYR	A	144.88	15.65	2	C	148.57	15.46	1
173	ASP	A	34.11	3.68	3	C	29.62	3.08	2
175	LEU	A	75.73	8.18	1	C	71.17	7.41	1
176	LYS	A	96.27	10.40		C	83.61	8.70	
177	THR	A	8.64	0.93	1	C	6.62	0.69	1
178	GLU	A	3.51	0.38		C	4.02	0.42	
179	LYS	A	40.84	4.41		C	29.01	3.02	
256	TYR	A	1.86	0.20		-	-	-	
264	PHE	-	-	-		C	3.13	0.33	

a: ΔASA buried accessible surface area
b: HB hydrogen bond

Table S2. AR_3a Residues on Molecules B and D Involved in the Interaction with mAPH Molecules A and C, Respectively

Residue#	Residue	Chain	Interface ΔASA[a]	% Interface ΔASA[a]	HB[b]	Chain	Interface ΔASA[a]	% Interface ΔASA[a]	HB[b]
16	LYS	B	38.44	4.54	1	D	27.55	2.90	
20	GLU	B	30.80	3.64		D	39.95	4.20	
23	ARG	B	107.87	12.75	2	D	115.72	12.17	2
43	ASN	-	-	-		D	1.56	0.16	
44	ASP	B	4.39	0.52		D	31.59	3.32	
45	TRP	B	107.79	12.74	1	D	73.90	7.77	
46	PHE	B	91.92	10.87		D	151.67	15.96	1
47	GLY	-	-	-		D	1.31	0.14	
48	ILE	B	36.95	4.37		D	34.35	3.61	
52	HIS	B	2.80	0.33		D	2.62	0.28	
53	LEU	B	11.12	1.31		D	17.03	1.79	
56	ASN	B	50.62	5.98	1	D	51.71	5.44	
57	ASN	B	29.80	3.52		D	27.23	2.86	
77	ASP	B	14.94	1.77	1	D	11.52	1.21	1
78	LYS	B	31.96	3.78	2	D	80.35	8.45	
79	SER	B	46.11	5.45	1	D	53.68	5.65	1
81	TRP	B	53.73	6.35		D	43.55	4.58	
86	LEU	B	19.71	2.33		D	19.44	2.05	
89	TYR	B	57.78	6.83	1	D	63.13	6.64	1
90	ARG	B	53.17	6.28	2	D	49.25	5.18	1
92	HIS	B	7.95	0.94		D	1.95	0.21	
110	ASP	B	2.07	0.25		-	-	-	
111	TYR	B	9.69	1.15		-	-	-	
112	GLN	B	10.46	1.24		D	18.67	1.96	
114	TYR	B	7.92	0.94		-	-	-	
119	LEU	B	13.21	1.56		D	8.91	0.94	
122	GLU	B	4.78	0.57		D	12.57	1.32	
123	ASP	-	-	-		D	11.32	1.19	

a: ΔASA buried accessible surface area
b: HB hydrogen bon

Structure, Vol. 13, 1089–1095, August, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.07.001

Previews

Allostery Trumps Antibiotic Resistance

In this issue of *Structure*, Kohl and colleagues report on the structure of an antibiotic resistance kinase in complex with an inhibitor selected from an ankyrin repeat library that binds the enzyme with high affinity and reverses antibiotic resistance in vivo (Kohl et al., 2005).

Kinases are ubiquitous modulators of protein function, modifiers of small molecules and nucleic acids, key catalysts in biotransformations, and essential components in cellular networks that sense and respond to external and internal stimuli. The superfamily of Ser/Thr/Tyr kinases has in particular garnered tremendous attention from the research community as a result of their vital roles in cellular signal transduction. Here protein and the related lipid kinases play key switching roles in cell death and mitogenesis, controlling the cell cycle and associated phenomena. As a result, this superfamily of enzymes figures prominently in associated diseases of cell growth and regulation such as cancer. Modulation of kinase activity therefore has been targeted by members of the scientific community and the pharmaceutical sector interested in identifying new small-molecule anticancer agents.

Members of the protein kinase superfamily share overall structural homology, with an N-terminal β sheet domain and a C-terminal α-helical domain. All kinases bind ATP with micromolar dissociation constants in a structurally conserved deep binding pocket. The generally more shallow phosphate-accepting substrate binding site, which is located primarily in the structurally divergent C-terminal domain, is difficult to block with nonpeptide inhibitors. As a consequence, the focus of most kinase-targeted drug discovery campaigns has been on the ATP binding pocket, which is readily amenable to small-molecule binding and for which ample precedent is available. The challenge inherent in this strategy is the ubiquity of cellular protein kinases and the resultant off-target interactions of kinase inhibitors. Approaches that identify new binding strategies to inhibit kinases, both as in vivo tools to probe kinase function and as drug leads, are therefore highly sought.

In this issue, Kohl et al. (2005) report the structure of an inhibited kinase issuing from a new kinase inhibitor binding strategy (Amstutz et al., 2005). The authors have taken advantage of an antibiotic resistance kinase, APH(3′)-IIIa, that phosphorylates aminoglycoside antibiotics. This kinase shares structural homology with protein kinases (Hon et al., 1997), has detectable protein kinase activity (Daigle et al., 1999), and can be inhibited by inhibitors of protein kinases that target the ATP binding site (Daigle et al., 1997). The antibiotic resistance phenotype conferred by APH(3′)-IIIa facilitates inhibitor screening, and therefore this enzyme is gaining popularity as a model for research on protein kinase screening and inhibition (e.g., Zhang et al., 2005).

The inhibitor of APH(3′)-IIIa identified by Kohl and colleagues is an ankyrin repeat protein selected for specific binding to the kinase. Ankyrin repeats are common elements incorporated into modular proteins that participate in protein-protein interactions (Mosavi et al., 2004). The ankyrin repeat element (AR) consists of a helix-loop-helix motif connected to another repeat through a variable loop or β-hairpin. This helix-loop-helix-loop/β-hairpin motif forms the core element in the AR structure, and a single protein can include anywhere from 2 to over 30 tandem AR motifs. These AR domains form extended surfaces suitable for protein-protein interactions. As a result, the motif appears in a number of biochemical contexts, including cell surface proteins and regulatory proteins.

Plückthun and colleagues have developed the AR repeat for engineering of protein binding proteins (Binz et al., 2004). Using a modular approach consisting of N- and C-terminal caps sandwiching AR cassettes, they have designed and constructed protein libraries that display random amino acids on one surface of the protein (Figure 1A). These libraries are then used in ribosome-display selection to identify AR binders of bait proteins. Using APH(3′)-IIIa in this mode, the group identified several members of a randomized library that bound to the kinase with low dissociation constants (nM) and inhibited the in vitro aminoglycoside antibiotic phosphorylation activity (Amstutz et al., 2005). Furthermore, coexpression of some of these AR inhibitors in *E. coli* expressing APH(3′)-IIIa reversed the antibiotic resistance phenotype, thereby demonstrating that these selected AR proteins have both potent in vitro and in vivo activity. The mode of interaction of these AR proteins with APH(3)-IIIa was, however, unknown until now.

The report in this issue by Kohl et al. (2005) addresses the question of how AR inhibition of aminoglycoside kinase occurs through determination of the X-ray structure of APH(3′)-IIIa in complex with one of the more potent AR proteins selected in the screen, AR_3a. The costructure reveals that AR_3a binds to the C-terminal helical aminoglycoside binding domain of APH(3′)-IIIa. Remarkably, the AR protein does not bind in a competitive fashion with antibiotic in the aminoglycoside binding pocket, but rather to an orthogonal region of the protein. This nonactive site interaction results in trapping of an inactive conformation of the enzyme. Binding of aminoglycoside substrates to APH(3′)-IIIa is associated with movement of helices in the C-terminal domain and formation of an obligate ionic interaction between the antibiotic substrate and the carboxylate of the C-terminal residue Phe264 (Thompson et al., 1999). The interaction of APH(3′)-IIIa and AR_3a results in a conformational change that distorts the aminoglycoside binding pocket through movements of key α helices, leading to conformation(s) of the C-terminal domain that are unproductive for antibiotic bind-

Structure
1090

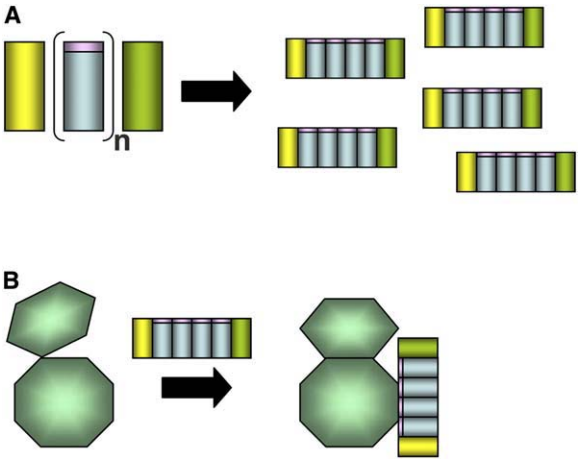


Figure 1. AK Repeat Libraries and Inhibition of APH(3′)-IIIa

(A) The AK repeat library consists of N- and C-terminal caps (in yellow and green, respectively) and a series of AK modules (blue) with a variable region (red) that presents on one side of the folded protein.

(B) Interaction of AR_3a with APH(3′)-IIIa (green) results in a conformational change that alters the aminoglycoside binding site and prevents productive binding of antibiotic for detoxification.

ing (Figure 1B). This structural rearrangement also impacts the positioning of the free carboxylate of Phe264 such that it is no longer available to stabilize the antibiotic in the binding pocket. The result of this interaction is powerful inhibition of antibiotic kinase activity and reversal of resistance in vivo.

This work mirrors the interactions of regulatory proteins with cognate protein kinases that are common in many eukaryotic cells to control kinase activity. The AR library approach demonstrates that this paradigm can be harnessed in the discovery of novel reagents to modulate cellular enzyme activity, which complements other strategies such as gene inactivation and RNAi. Furthermore, this research demonstrates that novel inhibitory modalities with biological impact can be discovered using the modular AR library approach coupled with structure determination. In this specific example, a new inhibitor binding mode has been uncovered for an important antibiotic resistance enzyme. This type of interaction can now serve as a model for the design of small-molecule inhibitors that inhibit APH(3′)-IIIa in a similar fashion, which could find clinical utility in blocking antibiotic resistance. Furthermore, the affinity of the AR_3a inhibitor for the enzyme also provides an opportunity to develop displacement assays that could be used to discover these molecules in a random screen. This general approach should be amenable to several enzyme classes, especially where a powerful biological readout (e.g., drug resistance in this case) can serve as primary screen.

Gerard D. Wright
Antimicrobial Research Centre
Department of Biochemistry and Biomedical Sciences
McMaster University
1200 Main Street West
Hamilton, Ontario, L8N 3Z5
Canada

Selected Reading

Amstutz, P., Binz, H.K., Parizek, P., Stumpp, M.T., Kohl, A., Grütter, M.G., Forrer, P., and Plückthun, A. (2005). J. Biol. Chem. *280*, 24715–24722.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. (2004). Nat. Biotechnol. *22*, 575–582.

Daigle, D.M., McKay, G.A., and Wright, G.D. (1997). J. Biol. Chem. *272*, 24755–24758.

Daigle, D.M., McKay, G.A., Thompson, P.R., and Wright, G.D. (1999). Chem. Biol. *6*, 11–18.

Hon, W.C., McKay, G.A., Thompson, P.R., Sweet, R.M., Yang, D.S.C., Wright, G.D., and Berghuis, A.M. (1997). Cell *89*, 887–895.

Kohl, A., Amstutz, P., Parizek, P., Binz, H.K., Briand, C., Capitani, G., Forrer, P., Plückthun, A., and Grütter, M.G. (2005). Structure *13*, this issue, 1131–1141.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.-Y. (2004). Protein Sci. *13*, 1435–1448.

Thompson, P.R., Schwartzenhauer, J., Hughes, D.W., Berghuis, A.M., and Wright, G.D. (1999). J. Biol. Chem. *274*, 30697–30706.

Zhang, C., Kenski, D.M., Paulson, J.L., Bonshtien, A., Sessa, G., Cross, J.V., Templeton, D.J., and Shokat, K.M. (2005). Nat. Methods *2*, 435–441.

Chapter 6

Rapid Selection of Highly Specific JNK2 Binding Ankyrin Repeat Proteins

Contents

Rapid Selection of Highly Specific MAP Kinase-Binders from Designed Ankyrin Repeat Protein Libraries

Patrick Amstutz [1], Holger Koch [1], Hans Kaspar Binz, Patrik Forrer, and Andreas Plückthun *

We here describe the rapid selection of highly specific c-jun N-terminal kinase 2 (JNK2) binders from a combinatorial library of designed ankyrin repeat (AR) proteins. A combined *in vitro*/*in vivo* selection approach, based on ribosome display and the protein fragment complementation assay (PCA) yielded a large number of different binders, fully functional in the bacterial cytoplasm. Only one round of ribosome display with subsequent small scale PCA selection was necessary to isolate a first specific binder. After two rounds of ribosome display selection, virtually all AR proteins identified by PCA showed specific binding, with affinities in the low nM range. The selected AR proteins specifically bound to JNK2, while they did not interact with the 81% identical JNK1. Taken together, the combination of ribosome display and PCA selections allowed the identification of large pools of binders at unparalleled speed. Moreover, quite in contrast to other intrabody selection protocols, our remarkably high selection efficiency demonstrates the potency of this alternative selection approach.

Keywords: ribosome display, protein fragment complementation assay, MAP kinase, JNK2, ankyrin repeat proteins, intrabody

Abbreviations used: AR, ankyrin repeat; pD, major surface protein of bacteriophage lambda; ELISA, enzyme-linked immunosorbent assay; JNK, c-Jun N-terminal kinase; MAPK, mitogen-activated protein kinase; mDHFR, murine dihydrofolate reductase; PCA, protein fragment complementation assay; SPR, surface plasmon resonance; TMP, trimethoprim.

Biochemisches Institut der Universität Zürich, Winterthurerstr. 190,
CH-8057 Zürich, Switzerland
[1] both authors contributed equally to this study
*Corresponding author
Tel. +41 1 635 55 70, Fax +41 1 635 57 12
E-mail address of the corresponding author: plueckthun@bioc.unizh.ch

Introduction

Mitogen activated protein kinases (MAPKs) are key factors in cellular signal transduction and are involved in a whole range of diseases, including cancer and inflammation [1]. The investigation of specific MAPKs is challenging, as these enzymes are highly homologous, with highly conserved active sites. This problem becomes even more pronounced for distinguishing members within the different subfamilies. For example, the c-Jun NH_2-terminal kinase (JNK) subfamily (Kallunki et al., 1994) consists of three members, JNK1, JNK2 and JNK3, each with different splice variants. While JNK1 and JNK2 are expressed ubiquitously, JNK3 expression is mostly restricted to the brain, heart and testis [2]. These kinases are responsible for the activation of cellular responses to extracellular signals (Davis, 2000). JNK-dependent signal transduction is important in a wide range of different processes, including cellular proliferation, oncogenic transformation or programmed cell death and many others. While JNK specific functions are known, assigning them to the different JNK isoforms is still difficult. For this purpose binding molecules which can reliably differentiate between these isoforms, are desired. Antibodies, the most commonly used binding molecules, are not ideal to target MAPKs, as these enzyme are located within the cell and antibodies tend not to sustain such a reducing environment, as their stability relies on disulfide bonds [3]. While different successful attempts have been reported to select antibodies, which are functional inside the cell, the number of selected and characterized binders is usually limited [4; 5]. These results might be explained by the selection pressure, which is not only specific binding but also stability under reducing conditions. Although considerable scientific effort has focused on the construction of stable antibody libraries, which are adapted to reducing selection conditions [6], the intracellular selection of specific antibodies still remains challenging. To overcome the limitations of antibodies and their derivatives, a variety of novel scaffolds for the generation of antibody-like binding molecules, possessing favorable biophysical properties, has been developed [7]. Designed ankyrin repeat (AR) proteins represent such a novel scaffold [8; 9]. These proteins are built from single repeat modules of 33 amino acids with fixed framework residues (27 amino acids) and randomized potential interaction residues (6 amino acids), which stack together to form the AR protein. This modular architecture is ideal for the assembly of very large combinatorial AR protein libraries. The library used for the selection experiments described here was built from three randomized AR modules, flanked by an N- and a C-terminal capping repeat shielding the hydrophobic core (N3C). While the theoretical diversity of such a library is 3.79×10^{23}, the experimentally sampled one had a diversity of greater than 10^{10}. As the number of modules in one designed AR protein can be adjusted [9], the target binging surface may be adapted as desired and is not fixed in dimensions as it is for other scaffolds, including antibodies.

It has been shown that the designed AR proteins match all requirements for functional intracellular applications: these molecules are very well expressed in the bacterial cytoplasm, they are stable and cysteine-free [10]. Moreover, specific binders with high affinity and specificity can be selected from designed AR protein libraries [11]. While the selection of specific binders from other novel scaffold libraries has also been reported [12], their intracellular applications are scarce [13]. In contrast, specific binders and inhibitors of a prokaryotic kinase, which are fully functional inside the cell, have been selected from the AR protein libraries [14].

We here report and analyze the selection of JNK2-specific AR proteins. The combination of ribosome display, a complete *in vitro* selection technology, and protein fragment complementation assay (PCA), an *in vivo* selection technology, allowed the rapid selection of large pools of intracellularly active JNK2 specific binders.

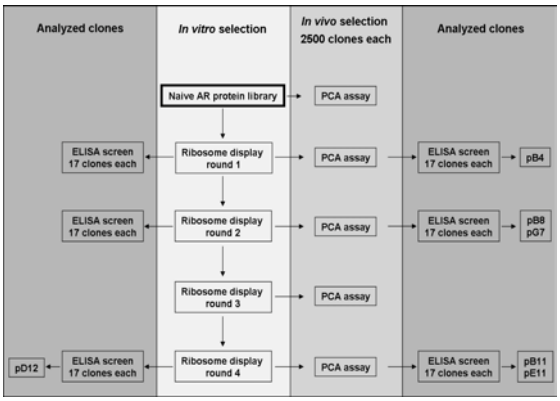


Figure 1. Schematic representation of the selection and analysis strategy.

Results

The goal of this project was the selection of AR proteins, which are able to discriminate between JNK2 and JNK1, to monitor the selection process and to test the functionality of the selected AR proteins *in vitro* as well as *in vivo*. For initial selections of binding molecules from large AR protein libraries, ribosome display was applied. As this complete *in vitro* selection technology circumvents any transformation step which would otherwise limit the applicable library size, it is ideally suited to handle large libraries (diversity $\geq 10^{10}$). PCA, a bacterial intracellular selection technology, based on the reconstitution of DHFR activity by two binding partners [15], was subsequently applied to analyze the ribosome-display generated pools of binders (Fig. 1), to further enrich these binders and to demonstrate their activity under reducing conditions. Selected AR proteins were further examined for their target binding properties with respect to affinity and specificity in ELISA and SPR experiments.

In vitro and in vivo selection of JNK2-binding AR proteins

For the selection of JNK2-specific binders from an N3C AR protein library, four standard ribosome-display selection rounds against immobilized JNK2 were performed [16]. A pool of DNA fragments encoding AR proteins (approximately 2500 individual clones) from the naïve starting library and after each ribosome-display selection round was inserted into the PCA selection plasmid as mDHFRI fusion. In parallel experiments, co-transformation of these pools with the JNK2-mDHFRII-bait-plasmid allowed determining the bacterial survival under selective conditions in dependence of the respective ribosome-display selection round (Fig. 2). As expected, nearly no bacterial growth resulted from co-transformation of the bait-plasmid with the naïve AR protein library. The pool after the first round of ribosome display also gave almost no growth under selective PCA conditions. In contrast, co-transformations of the AR protein pools corresponding to ribosome-display selection rounds 2, 3, and 4 gave rise to a large number of bacterial survivors, which increased with the selection round number (Fig. 2). As a control, the individual pools were also plated under non-selective conditions, where no difference in bacterial cell growth was observed, indicating that all pools had been treated equally (data not shown). These results suggest that two rounds of ribosome display are sufficient for selection of binders.

ELISA screen to identify JNK2-specific AR proteins

To elucidate whether the increased bacterial growth, observed in the *in vivo* PCA selection, did indeed result from an enrichment of JNK2-specific AR proteins, the binding of the selected AR proteins to JNK2 was tested in ELISA screens. Seventeen clones from each of the selection rounds 1, 2, and 4, obtained either with or without additional PCA selection were randomly picked. Crude *E. coli* extracts from small-scale expression cultures were used to compare the binding of the selected clones either to JNK2 or, as a control, to the bacteriophage lambda protein D (pD) in ELISA experiments (Fig. 3).

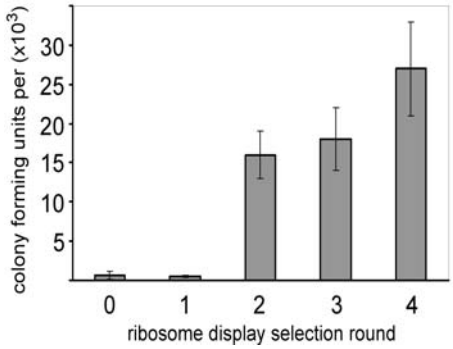


Figure 2. Number of colonies obtained by *in vivo* PCA selection with different AR protein library pools.
The naïve AR protein library and the AR protein pools obtained after each round of ribosome-display selection was cloned as N-terminal mDHFRI fusion (see Material and Methods) and co-transformed in *E. coli* with the JNK2-mDHFRII antigen construct under selective conditions. A comparable level of bacterial growth under selective conditions resulted from co-transformation of the naïve AR protein library or the AR protein pool of the 1st ribosome-display selection round with the JNK2-DHFRII encoding plasmid. In contrast, increasing numbers of colonies are observed from co-transformation of the AR protein pools of ribosome-display selection round 2, 3, and 4. Under non-selective growth conditions, a similar number of colonies were obtained in all cases.

Of the AR proteins, which were exclusively selected by ribosome display, no JNK2-specific AR protein was found in the pool from selection round 1, whereas approximately 30% of the AR proteins of both the selection rounds 2 and 4 showed JNK2 binding. In comparison, when a PCA selection was carried out after the first ribosome-display selection round, one JNK2 binder was obtained. Furthermore, approximately 80% of the AR proteins of round 2 and nearly 100% of the AR proteins of round 4 showed binding to JNK2, when the ribosome display rounds were followed by PCA selection. It seems that ribosome display is afflicted with an intrinsic background, which cannot be reduced by further selection rounds. Thus, the expression plasmids which were obtained by ligation of the ribosome-display product, encode a constant percentage of non-binders. While this was almost 70% in the present experiments, values are as low as 20% have also been observed [11]. In contrast, when the same pools are sieved by PCA selection, already after the second round of ribosome display, the majority of clones are specific binders. PCA alone, on the other hand, suffers from background problems, if selection is carried out with naïve libraries [5]. Thus, the combination of both methods allows an efficient analysis of binders after very few rounds, when there is still significant diversity.

Sequencing of the JNK2-specific AR proteins
From the group of JNK2-specific AR proteins, which were identified in the initial ELISA screen, 10 AR proteins resulting from exclusive ribosome-display selection and 24 AR proteins obtained by the additional PCA selection were chosen for DNA sequencing. From the binders which were selected only by ribosome display none was found multiple times. In contrast, the JNK2 binder pB4, which was isolated from the PCA selection following ribosome-display round 1, was found three times (pB4, pC4, and pD4). Another binder was found twice, once in round 2 (pF7) and once in round 4 (pA12) from the PCA selection following ribosome display. Furthermore, this particular binder (pF7, pA12) was also found in the pool of only ribosome-display selected AR proteins (rD12). Another binder was found in both groups of ribosome-display selected (rC6) and PCA enriched (pE8) AR proteins (for sequences of the selected AR proteins see suppl. info.).
Of the non-binder population, 13 clones without and 13 clones with additional PCA selection were sequenced. Thereby, no deletions, premature stop codons or frame shifts were detected. However, a substantial variation of sizes, ranging from one to five internal repeats was found. This finding might be explained by recombination events during RT-PCR.

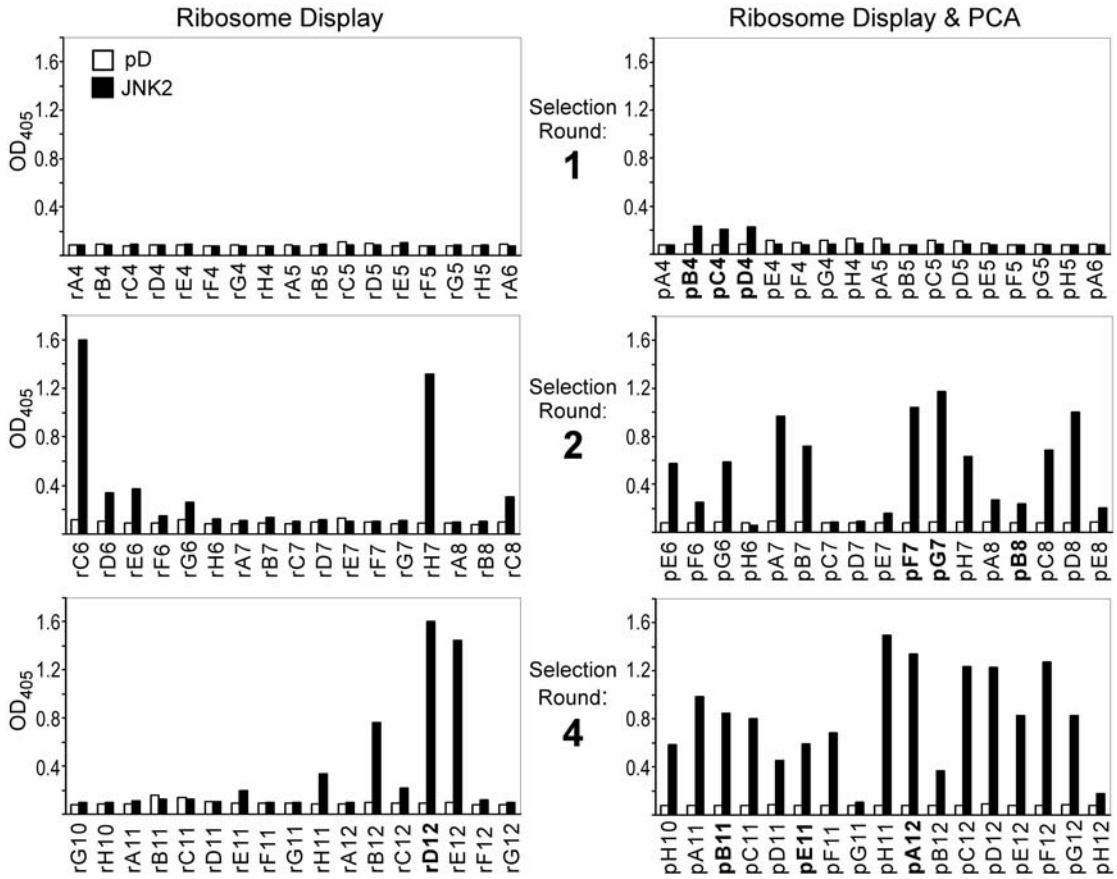


Figure 3. ELISA screening for the identification of JNK2-specific AR proteins.
The JNK2-binding of randomly picked AR proteins of different ribosome-display selection rounds (with or without additional PCA selection step) are compared. Crude cell extracts of *E. coli* expressing the AR proteins as mDHFRI fusions were applied to wells coated with JNK2 or pD (control). No JNK2-specific AR protein was identified from the group of potential binders resulting from selection round 1 of the ribosome display selection, whereas three JNK2-specific AR proteins (identical clones) were found in round 1 of the additional PCA selection. From the groups of AR proteins, originating from ribosome-display selection round 2 and 4, approximately 30% possess JNK2-specificity. Again, many more JNK2-specific AR proteins were identified from the corresponding selection rounds after additional PCA selection: approximately 80% of the AR proteins of round 2, and nearly all AR proteins of round 4 show specific binding to JNK2. All the binders, which were chosen for further characterization, are marked bold (pB4=pC4=pD4; rD12=pF7=pA12). In summary, the performance of the additional PCA screen leads to a much stronger enrichment of JNK2-specific AR proteins in all evaluated selection rounds.

Also, when monitoring the outcome of the RT-PCR after a ribosome-display selection round by agarose gel electrophoresis, a certain proportion of the AR proteins were found to be shortened by one AR module (N2C instead of N3C). This shortening was found for 10% of the ribosome-display selected AR proteins, but, surprisingly, for 50% of the AR proteins resulting from the subsequent PCA screen. This observation might indicate a growth advantage of the N2C AR proteins over the N3C molecules in this context of the PCA selection system.

In conclusion, from 34 sequenced JNK2-binding AR proteins, 29 different AR proteins were

identified, indicating that the generated binder pool still holds substantial diversity.

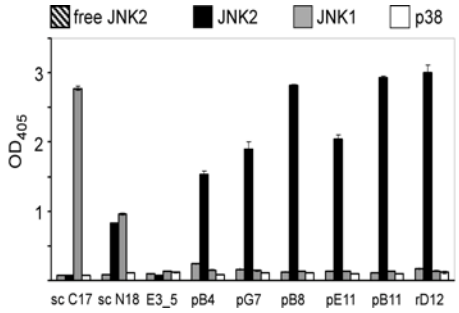


Figure 4. Cross-specificity comparison of selected AR proteins and commercially available JNK-specific antibodies by ELISA.

Binding of different selected AR proteins to JNK2, JNK1 and p38 is compared. In addition, the inhibition of JNK2-binding by free JNK2 is tested for each molecule. Two commercially available antibodies (sc C17 and sc N18, see Material and Methods) and one unselected AR protein (E3_5) served as controls. Purified AR proteins were applied at a concentration of 250 nM, whereas the purchased antibodies were diluted (1:500) according to the recommendation of the supplier. As the dissociation constant of the AR protein pB4 is higher, pB4 was applied at a concentration of 5 μM. The binding to immobilized JNK2 of all binders including the purchased antibodies was inhibited completely by a threefold excess of free JNK2. All selected AR proteins and the JNK1 antibody (sc C17) were highly specific to their respective antigen, while the JNK2 antibody (sc N18) does not discriminate between the JNK isoforms. The unselected AR protein (E3_5) does not bind to any of the evaluated antigens. Furthermore, none of the tested binding proteins show interaction with p38.

Specificity and affinity of selected AR proteins

Target binding of 6 JNK2-binding AR proteins was characterized by surface plasmon resonance (SPR) and by ELISA experiments. Binders from the different selection rounds, with and without additional PCA-selection step, were included (Fig. 3).

Kinetic SPR measurements using multiple concentrations of the selected AR proteins were used to determine their dissociation constants (Table1). While the K_D of pB4, the clone that resulted from the first selection round, was around 1 μM, all other binders had low nM affinities. Interestingly, the affinity of binders from selection round 2, can equal the ones from selection round 4. Binders generated with or without additional PCA selection seemed to bind with similar affinities (Table1). As PCA yielded binders possessing affinities ranging from the nM to the μM range, it seems that the selection system can tolerate a broad spectrum of affinities, all resulting in bacterial growth.

ELISA experiments were used to investigate the specificity of the JNK2 binders. The binding to JNK2 was compared to that to JNK1 and p38. While JNK2 and JNK1 share 81% identity and 86% similarity, JNK2 and p38 still share 51% identity and 59% similarity on the amino acid level in our format. For comparison, two commercially available antibodies (see Material and Methods), one binding to JNK1, the other to JNK2, were tested in parallel. All AR proteins and the JNK1 antibody were highly specific to their cognate antigen, while the JNK2 antibody seemed to not discriminate between JNK1 and JNK2 (Fig. 4). None of the binding proteins interacted with p38. We also tested whether the binding to immobilized JNK2 could be inhibited by free JNK2. The binding to immobilized JNK2 of all binders including the antibodies was inhibited completely by a threefold excess of free JNK2 over the binder (Fig. 4).

Clone	Round	Selection	Size	k_{on} (1/Ms)	k_{off} (M)	K_D (nM)
pB4	1	RD&PCA	N3C	$(4.14 \pm 2.15)\cdot 10^{4}$	0.0342 ± 0.0041	914 ± 568
pB8	2	RD&PCA	N2C	$(1.42 \pm 0.63)\cdot 10^{5}$	$(9.24 \pm 2.93)\cdot 10^{-4}$	6.5 ± 4.3
pG7	2	RD&PCA	N3C	$(3.5 \pm 1.31)\cdot 10^{5}$	$(4.0 \pm 0.57)\cdot 10^{-3}$	11.3 ± 5.9
pB11	4	RD&PCA	N3C	$(1.96 \pm 0.48)\cdot 10^{5}$	$(3.98 \pm 0.58)\cdot 10^{-3}$	22.0 ± 9.8
pE11	4	RD&PCA	N2C	$(4.15 \pm 1.47)\cdot 10^{5}$	$(9.57 \pm 0.73)\cdot 10^{-4}$	2.31 ± 1.0
rD12	4	RD	N2C	$(4.67 \pm 0.24)\cdot 10^{5}$	$(6.75 \pm 0.92)\cdot 10^{-4}$	1.44 ± 0.3

Table 1. Affinity of selected AR proteins

Discussion

We demonstrate here the selection of a large pool of intracellularly active AR proteins binding to JNK2 with very high specificities and affinities. For the generation of kinase-specific AR proteins a combined *in vitro* and *in vivo* selection was performed. Ribosome display was applied for the selection of JNK2 binders from a large combinatorial AR protein library. PCA was used to monitor the ribosome-display selection process, to further enrich specific AR proteins from the pools of binders generated by ribosome display and to test the functionality of the selected AR proteins under reducing conditions. The analysis of the binders from different selection rounds showed that the selection process of obtaining kinase-specific AR proteins was extremely rapid. Only two rounds of ribosome display were sufficient for enrichment of a whole variety of different JNK2-specific AR proteins. Almost all selected AR proteins possess high specificity and affinity. These results suggest that the combination of the AR protein library with both ribosome display and PCA selections is an ideal set-up to rapidly generate diverse pools of specific binders, fully functional under reducing conditions. These pools can then be used as starting points to screen for functionality for a whole range of different applications, including intracellular ones.

Targeting kinases – an issue of specificity

Up to date, around 518 human protein kinases (EPKs) are known [17]. Most of these are part of complex signalling cascades, and the different kinases often share very high homology. This homology complicates the selection of specific EPK binders. To address EPK function, often binders and inhibitors are used, which lack sufficient specificity, and thus the observed effects cannot be allocated correctly. All selected AR protein binders reported here were highly target kinase specific, binding to JNK2, while not recognizing JNK1, despite over 80% identity of the target proteins. This high specificity of designed AR proteins may be a useful in the quest to understand the function of individual kinases in more detail.

We believe that the high specificity of the selected AR proteins has two reasons: the potential binding epitope and the binding mode. Most peptides and small molecules used to date bind EPKs in grooves [18]. These grooves, including the active site, are often highly conserved, preventing specific targeting. As the randomized positions on the AR protein span a large surface the selected binders will interact with the target protein on exposed surface areas. These surface areas are more likely to be kinase-specific, and thus specificity is achieved more easily. From the crystal structures of AR protein library members, unselected or in complex with the target [10; 11], we know that the AR protein

behaves essentially like a rigid body. This might restrain the binding epitopes, but results in very high specificity, as the AR protein will not adapt to conformations of other proteins.

Selection process – rapid enrichment of many functional binders

The efficient selection of specific and stable binding molecules, which remain functional under reducing conditions, demands new selection technologies. Here we present an alternative approach, in which PCA was used as read-out for antigen-specific AR proteins previously selected by ribosome display. By combination of ribosome display and PCA, not only a very rapid enrichment of antigen-specific AR proteins was revealed, but also all selected AR proteins could simultaneously be tested for functionality under reducing conditions. The monitored window of 2500 clones of every ribosome-display selection round that was applied to PCA was sufficient to yield a high variety of different binders. Already after one round of ribosome-display selection the first binder was found by PCA and after round two, the selection process was virtually complete. If the starting library diversity indeed was in the range of 10^{10}, this equals an enrichment factor of approximately 10^6, a very high number for ribosome-display selections [19]. The diversity of the selected binders was remarkably high, as from 34 sequenced clones, only three clones were found multiple times, while 29 where different. Similar antibody-based selections with ribosome display, including five to six selection rounds (in the absence of PCA), tended to converge to a small number of high-affinity binders [19]. The high diversity of selected AR protein binders is especially attractive, if not only simple binding, but a specific function is desired. Here the pools of binders might serve a starting point for functional screens, as recently shown for a prokaryotic kinase [14].

This rapid enrichment of diverse pools of different binders might be explained by the high quality of the designed AR protein library. We believe that this scaffold, with its highly beneficial biophysical properties, is ideal for library generation. Due to its inherent stability the designed AR domain fold will tolerate a very large diversity in the potential interaction surface, leading to a very large percentage of functional library members. If more proteins are functional, fewer will be aggregated and display exposed hydrophobic patches. This high percentage of folded proteins will reduce the problem of unspecific binding during ribosome-display selections. Furthermore, a large and diverse population of active library members increases the probability of successfully selecting many different binders [20].

Selection method – ribosome display and PCA

Successful selections from large protein libraries demand potent selection technologies. Ribosome display, which works completely *in vitro*, allows the handling of very large libraries, as no transformation step limits the applicable library size [16]. Indeed the combination of the AR protein libraries with ribosome display rapidly yielded several high-affinity binders. Nevertheless an intrinsic background of non-target binders was observed, remaining constant from ribosome-display selection round two on. The origin of this background was not determined. To eliminate this background as well as to analyze the selection process, PCA selections were performed. In the PCA experiments the background was completely eliminated. Nevertheless, a direct PCA selection from naïve libraries seems still difficult, as this selection yielded survivors not linked to a specific binding event. Generally, PCA selections from naïve protein libraries, while able to select specific binders, seem to suffer from background problems, also for antibody-based selections [5].

We believe that ribosome display for selecting large pools of binders in combination with PCA for sieving these pools is an ideal combination for rapid identification of many different binders. This combination of *in vitro* selection and *in vivo* screen has also been described for the selection of intracellulary active antibody fragments [4; 6].

Usually phage display had been applied for the initial selection of binders and then yeast two-hybrid systems had been applied to select the intracellularly active ones. Thereby, only very stable antibodies sustained the reducing intracellular environment, and many binders might be lost due to insufficient stability. One strategy to overcome this problem is the use of antibody libraries based on highly stable frameworks [21]. AR proteins are *per se* stable under reducing conditions and therefore, probably no binders will be lost in the second *in vivo* selection step.

Material and Methods

Molecular biology: Unless stated otherwise, all experiments were performed according to the protocols of Sambrook *et al.* [22]. Enzymes and buffers were purchased from New England Biolabs (Beverly, MA, USA) or Fermentas (Vilnius, Lithuania). Oligonucleotides were obtained from Microsynth (Balgach, Switzerland).

***Antigen expression and purification*:** For expression of the antigen constructs the vectors pAT222 (GenBank accession **AY327137**), pAT222_p38 and pAT222_JNK2 [11], as well as pAT222_JNK1 were used. pAT222_JNK1 was constructed by inserting the PCR amplified JNK1 open reading frame (oligonucleotides jnk1f 5'-TTC CGC GGA TCC GGT ACC TCC CGT AGC AAG CGT GAC AAC -3' and jnk1r 5'- AAA CCC AAG CTT GCT GCA CCT GTG C -3'; template pAT58, [23] into pAT222, using the restriction enzymes *Bam*HI and *Hin*dIII. Expression from the vectors pAT222, pAT222_p38, pAT222_JNK1 and pAT222_JNK2 yields a fusion protein comprising an N-terminal avi-tag for *in vivo* biotinylation, bacteriophage lambda protein D (pD), followed by the target molecule (or nothing in the case of pAT222), and a C-terminal His-tag for purification (avi-pD-target-His_6). Alternatively, for expression of His_6-tagged JNK2 without any further fusion partner, the vector pQE30_JNK2 was used. pQE30_JNK2 was constructed by inserting the *Bam*HI/*Hin*dIII JNK2 fragment of pAT222_JNK2 into pQE30 (QIAgen, Hilden, Germany). Expression, biotinylation and purification of His_6-tagged proteins pD, pD_JNK1, pD_p38, pD_JNK2, and non biotinylated JNK2 were performed as described [11]. Biotinylation was quantified by using ELISA and SDS-PAGE followed by blotting with a streptavidin-alkaline phosphatase conjugate (Roche, Basel, Switzerland).

***In vitro ribosome display selection*:** The PCR amplified AR protein libraries were transcribed and 4 selection rounds were performed as described [11]. After each ribosome-display selection round the RT-PCR product of the encoding AR protein pool was analyzed by agarose gel-electrophoresis. To prevent an enrichment of N2C molecules, the N3C pool of each selection round was gel purified prior to re-amplification of the corresponding DNA. The DNA of each selection round was finally inserted in the PCA selection plasmid and cloned by transformation of *E. coli* XL10 (Stratagene, La Jolla, CA, USA).

In vivo selection using PCA: The RT-PCR products of each ribosome-display selection round, as well as the naïve AR protein library, were separately amplified by PCR using the oligonucleotides Ank-Sph-FOR (5'- GCT CAG GCA TGC TTG ACC TGG GTA AGA AAC TGC -3') and Ank-Eco-BACK (5'- GCT GCA GAA TTC TTG CAG GAT TTC AGC CAG G -3') and inserted via *Sph*I and *Eco*RI into the PCA selection plasmid pHK46 [5]. Transformation of *E. coli* XL10 with the ligation products yielded approximately 2500 colonies for each of the five AR protein pools (four ribosome-display selection rounds and the naïve AR protein library) and virtually no background of re-ligated vector was observed. The colonies of each transformation were pooled separately and the plasmids were isolated.

For performance of the *in vivo* PCA selection electrocompetent *E. coli* BL21/pRep4 (Qiagen, Hilden, Germany) cells (transformation efficiency

$\geq 3 \times 10^8$/μg DNA) were co-transformed in parallel with 100 ng of each AR protein pool-mDHFRI and JNK2-mDHFRII fusion constructs. After 1 h incubation at 37°C the cells were washed with M9 minimal medium and plated on selective M9 minimal media supplemented with 50 μg/ml kanamycin, 100 μg/ml ampicillin, 10 μg/ml chloramphenicol, 2 μg/ml TMP, 100 μM IPTG, and 5 % (w/v) of casamino acids (Difco, Detroit, MI, USA) and incubated at 25°C for 72 h. In parallel to the selective co-transformation, *E. coli* BL21/pRep4 was transformed with 100 ng of each AR protein pool-mDHFRI fusion constructs and incubated overnight at 37°C on standard LB plates (no selection pressure).

***ELISA screening to identify specific JNK2-binding AR proteins*:** seventeen clones each, from ribosome-display selection rounds 1, 2, and 4, with or without additional PCA selection step (Fig. 1, Fig. 2), were randomly picked and cultured overnight in 1.2 ml LB (LB, 1% glucose, 100 μg/ml ampicillin, and 50 μg/ml kanamycin) in 96-deep-well plates (Abgene, Surray, UK). On the next day, 300 μl each of these pre-cultures were added to 1200 μl LB medium, and incubated at 37°C. After 1 h incubation, expression of the AR protein-mDHFRI fusions was induced by 0.5 mM IPTG, followed by incubation for 3 hours at 37°C. The cultures were harvested by centrifugation and subsequently lysed by shaking for 10 min in 50 μl B-Per (Pierce, Rockford, IL, USA). After addition of 250 μl TBS_{500} (50 mM Tris-HCl, pH 7.8, 500 mM NaCl) the lysed cells were centrifuged again (4°C). One hundred fifty μl supernatant each were transferred to a MaxiSorp plate (Nunc, Roskilde, Denmark), coated with either pD_JNK2 or pD (for immobilization of pD or pD_JNK2, see Binz *et al.* (2004)). After incubation for 45 min at 4°C, AR protein-mDHFRI fusions were detected using an anti-$RGS(His)_6$-antibody as described [11].

Expression and purification of selected AR proteins: DNA encoding the selected AR proteins were PCR amplified using oligonucleotides Ank-Sph-FOR and Ank-Hind-BACK (5'- GAG GAT CCA AGC TTC TAT TAT TGC AGG ATT TCA GCC AGG TC -3') and inserted via *Sph*I and *Hind*III into the expression vector pQE32 (Qiagen), which is the PCA selection plasmid without the mDHFR fusion protein. Successful cloning was verified by DNA sequencing. Expression and purification of the AR proteins was carried out as described [10]. Purity of the samples was assessed by SDS-PAGE and concentrations were determined by UV absorbance at 280nm.

Cross-specificity of selected AR proteins: The biotinylated proteins JNK2, JNK1, and p38 were immobilized on neutravidin-coated MaxiSorp plates (Nunc, Roskilde, Denmark) as described [11]. One hundred μl of selected and purified AR proteins (250 nM) were incubated in TBST (50 mM Tris-HCl, pH 7.8, 150 mM NaCl, 0.05% Tween, and 0.5% BSA) in each well for 30 min at 4°C. In addition, one unselected AR protein (100 μl 250 nM E3_5, [9]) and two JNK antibodies (Santa Cruz Biotechnology, Santa Cruz, CA, USA) in a dilution of 1:500 (dilution recommended by the supplier) in TBST served as controls (JNK1 antibody C17/sc-474 and JNK2 antibody N18/sc-827). Furthermore, one aliquot of each binder (AR protein and antibodies) was additionally pre-incubated with free JNK2 (750 nM) prior to incubation with the immobilized antigen. Clone pB4 was applied at higher concentrations (5 μM purified AR protein; inhibited with 15 μM free JNK2). Binding of the selected AR proteins was analyzed by detection of their N-terminal $RGS(His)_6$ motif as described [11]. Binding of the purchased JNK antibodies was detected by a goat anti-rabbit IgG-AP antibody conjugate (100 μl/well, 1:3000, 45 min; Sigma (St. Louis, MO, USA)). After extensive washing, the bound AR proteins as well as the purchased JNK antibodies were visualized using a 4-nitrophenyl phosphate solution (3 mM 4-NPP, 50 mM $NaHCO_3$, and 50 mM $MgCl_2$ in H_2O; 100 μl/well). Absorbance was measured at a

wavelength of 405 nm. All measurements were performed in triplicates.

***Surface plasmon resonance (SPR)*:** SPR was measured using a BIAcore3000 instrument (BIAcore, Uppsala, Sweden). All measurements were done in HBS buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1mM DTT, 0.005% Tween 20) at a flow rate of 50 μl/min. Biotinylated pD_JNK2 fusion protein was immobilized (600 RU) on a SA chip (BIAcore). For the determination of kinetic data, the interactions were measured as follows: five minutes initial buffer flow, followed by a two-minute injection of AR protein in varying concentrations (1 nM to 200 nM) and a final off-rate measurement of 45 minutes with buffer flow. Clone pB4 was measured at higher concentrations (50 nM to 1 μM) using a shorter dissociation time (15 min). The signal of an uncoated reference cell was always subtracted from the measurements. The kinetic data of the interaction were evaluated with a global fit using BIAevaluation 3.1 (BIAcore).

References

1. Manning, A. M. & Davis, R. J. (2003). Targeting JNK for therapeutic benefit: from junk to gold? *Nat. Rev. Drug Discov.* **2**, 554-565.
2. Gupta, S., Barrett, T., Whitmarsh, A. J., Cavanagh, J., Sluss, H. K., Derijard, B. & Davis, R. J. (1996). Selective interaction of JNK protein kinase isoforms with transcription factors. *Embo J.* **15**, 2760-2770.
3. Biocca, S., Ruberti, F., Tafani, M., Pierandrei-Amaldi, P. & Cattaneo, A. (1995). Redox state of single chain Fv fragments targeted to the endoplasmic reticulum, cytosol and mitochondria. *Biotechnology (N Y)* **13**, 1110-1115.
4. Tanaka, T. & Rabbitts, T. H. (2003). Intrabodies based on intracellular capture frameworks that bind the RAS protein with high affinity and impair oncogenic transformation. *Embo J.* **22**, 1025-1035.
5. Koch, H., Gräfe, N., Schiess, R. & Plückthun, A. (submitted for publication). Development of the Protein Fragment Complementation Assay to Select Functional Intracellular Antibodies.
6. Visintin, M., Settanni, G., Maritan, A., Graziosi, S., Marks, J. D. & Cattaneo, A. (2002). The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies. *J. Mol. Biol.* **317**, 73-83.
7. Nygren, P. A. & Skerra, A. (2004). Binding proteins from alternative scaffolds. *J. Immunol. Methods* **290**, 3-28.
8. Forrer, P., Stumpp, M. T., Binz, H. K. & Plückthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539**, 2-6.
9. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *J. Mol. Biol.* **332**, 489-503.
10. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. U S A* **100**, 1700-1705.
11. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G. & Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575-582.
12. Mathonet, P. & Fastrez, J. (2004). Engineering of non-natural receptors. *Curr. Opin. Struct. Biol.* **14**, 505-511.
13. Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. (2002). Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. *Proc. Natl. Acad. Sci. U S A* **99**, 1253-1258.

14. Amstutz, P., Binz, H. K., Parizek, P., Stumpp, M. T., Kohl, A., Grütter, M. G., Forrer, P. & Plückthun, A. (submitted for publication). Intracellular Kinase Inhibitors from Combinatorial Libraries of Designed Ankyrin Repeat Proteins.
15. Mössner, E., Koch, H. & Plückthun, A. (2001). Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs. *J. Mol. Biol.* **308**, 115-122.
16. Hanes, J. & Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U S A* **94**, 4937-4942.
17. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* **298**, 1912-1934.
18. Noble, M. E., Endicott, J. A. & Johnson, L. N. (2004). Protein kinase inhibitors: insights into drug design from structure. *Science* 303, 1800-5.
19. Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287-1292.
20. Perelson, A. S. (1989). Immune network theory. *Immunol. Rev.* **110**, 5-36.
21. Desiderio, A., Franconi, R., Lopez, M., Villani, M. E., Viti, F., Chiaraluce, R., Consalvi, V., Neri, D. & Benvenuto, E. (2001). A semi-synthetic repertoire of intrinsically stable antibody fragments derived from a single-framework scaffold. *J. Mol. Biol.* **310**, 603-615.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular cloning: A laboratory manual, edn. 2*, (Cold Spring Harbor Press, New York).
23. Forrer, P. & Jaussi, R. (1998). High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage lambda head protein D. *Gene* **224**, 45-52.

Chapter 7

Binding Molecules to Diverse Targets

Contents

Selecting Designed Ankyrin Repeat Proteins for Co-Crystallisation of Membrane Proteins

Patrick Amstutz, Simon Jenni, H. Kaspar Binz, Jasmine Lozza, Nenad Ban & Andreas Plückthun

SecYEG is a protein complex in the membrane, which forms a protein conducting channel, allowing polypeptides to be transported across or integrated into membranes (Matlack et al., 1998). To understand this fundamental mechanism in molecular detail, a three-dimensional structure is needed (Breyton et al., 2002; van den Berg et al., 2004). The crystallization of this complex is not trivial as it spans the membrane with 15 transmembrane helices (SecY, 10; SecE, 3; SecG, 2).

The aim of this project was to select specific binders of secYEG from AR protein libraries. These binders could be used to stabilize distinct conformations of the membrane protein and enlarge its hydrophilic surface, thus aiding the crystallization process, as it has been showed for antibodies (Dutzler et al., 2002; Ostermeier et al., 1995).

First, selection experiments were performed to show that ribosome display as well as the AR proteins function in the presence of detergents. For this purpose, one round of ribosome display with a maltose binding protein binder (off7) was carried out, in the presence and absence of detergents (dodecyl maltoside (DDM): 0.03%). The presence of DDM seemed not to have any influence on the outcome of the selection round (Fig. 1).

Four rounds of ribosome display with the designed AR protein library were carried out on immobilized secYEG (as described in Binz et al., 2004) with 0.03% DDM in all buffers. SecYEG carried an N-terminal biotinylated avi-tag on secY for immobilisation and a N-terminal His-tag on secE for purification. Crude extract ELISAs were performed to screen for specific binders. Of around 200 screened clones from selection round four, 15 showed specific and 10 showed unspecific

binding (data not shown). The binders were analysed in more detail in ELISA experiments (Fig. 2). Even though specific binders were obtained, the project was stopped at this stage as the crystal structure of secYEG was published (van den Berg et al., 2004).

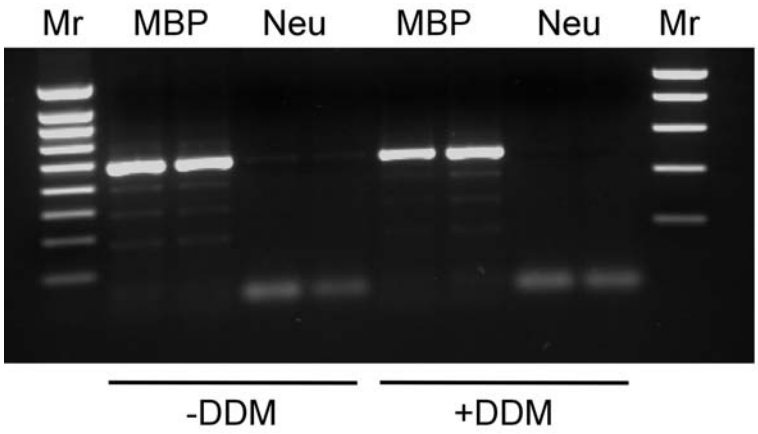


Figure 1. Ribosome display selection round with AR proteins in the presence of detergents. A standard ribosome display selection round was performed, displaying off7, a MBP binder (Binz et al., 2004). The round was done in duplicate and analysed at the level of PCR after RT, testing binding to MBP and background binding to neutravidin (Neu) in the presence and absence of 0.03% DDM. The outcome, specific binding signal if MBP was present and no signal if not, was not found to be detergent dependant.

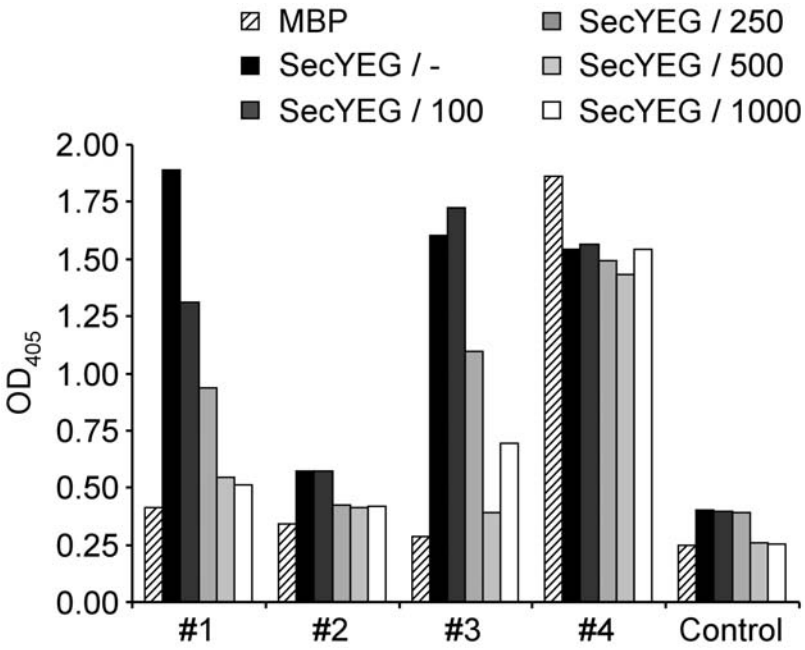


Figure 2. The binding of selected clones in ELISA experiments. The binding to secYEG was compared to binding to MBP and competition experiments with free secYEG (0 nM, 100 nM, 500 nM, 1000 nM) inhibiting the binding to immobilized secYEG were performed. The binders that had been selected against secYEG (#1, #2, #3) show no binding to MBP, while binding to immobilized secYEG. This interaction can be inhibited by addition of free secYEG. The unspecific binder (#4) gives rise to signal under all conditions, indicating that it sticks or binds to neutravidin or BSA, which were used for coating and blocking of the wells during ribosome-display selection and the ELISA.

Selection of Semliki Forrest Virus Inhibitors from Designed Ankyrin Repeat Protein Libraries

Stefan Deuber, Patrick Amstutz, H. Kaspar Binz, Andreas Plückthun & Jovan Pavlovic

Semliki Forrest Virus (SFV) is a positive stranded RNA virus. SFV is a member of the Alpha virus genus of the Togaviridae family, which replicates in the cytoplasm of the host cell. The viral replication machinery is encoded by its non-structural proteins 1-4 (2342 amino acids), which are synthesized as one polyprotein and subsequently cleaved (Kujala et al., 2001). The functions of the different NSPs are still under investigation. It is known that NSP1 is involved in RNA capping and mutations leading to inhibition of the guanylyltransferase activity are lethal to the virus (Ahola & Kääriäinen, 1995; Wang et al., 1996). NSP4 was shown to have RNA polymerase activity, whereas the functions of NSP3 are less clear. NSP2 has several different functions including NTPase, helicase and protease activity (Kujala et al., 1997).

The aim of this project was to select ankyrin repeat (AR) proteins as intracellular viral inhibitors. Inhibition of NSP function by an AR protein inside the cell should inhibit viral replication. In analogy to the selection of APH inhibitors, our plan of action was to use ribosome display to select binders of the NSPs from AR protein libraries and to use these pools of binders for *in vivo* screens detecting individuals protecting cells from SFV.

The initial experiments included panning against NSP2. Four rounds of standard ribosome display were performed (as described in Binz et al., 2004)). The pool of selected binders was cloned by ligation to pQE30. Single clones were analysed in crude extract ELISAs for binding specificity (Fig 1). In the meantime the production of all NSPs was improved and selections against NSP1, NSP2 and NSP4 are ongoing.

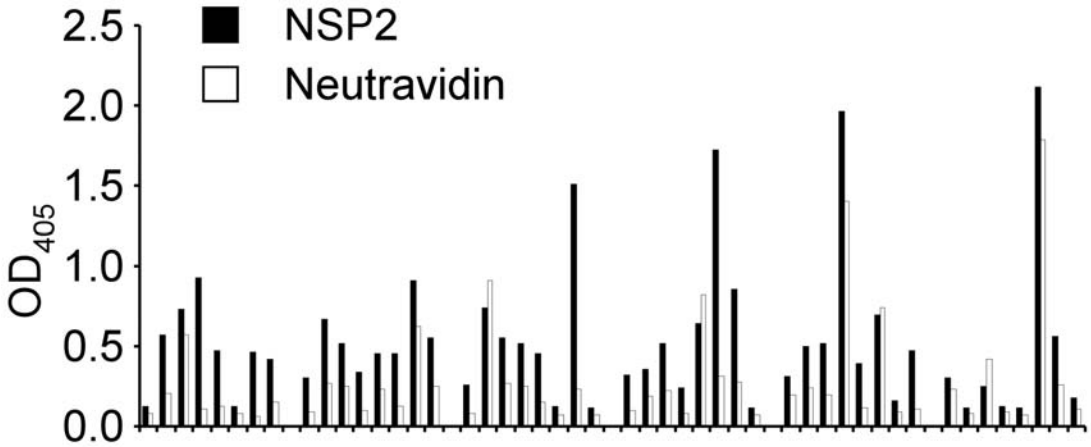


Figure 1. Crude Extract ELISA for the identification of NSP binders after four rounds of ribosome display selection. Of 48 analysed clones, four binders showed strong and specific binding (ratio: specific/unspecific binding >5), 19 showed specific binding (ratio: >2.5), 22 showed no binding and four were unspecific (ratio: ~1). We concluded that a pool of specific binders was selected, containing also unspecific ones.

Selection of Binders and Activity Modulators of AMPK from Designed Ankyrin Repeat Protein Libraries

Nadine Straumann, Patrick Amstutz, H. Kaspar Binz, Patrik Forrrer, Andreas Plückthun & Theo Wallimann

AMP-activated protein kinase (AMPK) is the metabolic master switch of the cell. AMPK senses the AMP level in the cell, not to be confused with cAMP, detecting stress to meet cellular energy demands (Kemp et al., 1999). Low AMP levels, as found in exercising muscle, activate AMPK. After activation, energy-producing processes, such as for example glucose uptake and metabolism, are started, and energy consuming ones, such as fatty acid synthesis, transcription or apoptosis, are stopped. This central role in energy sensing might suggest AMPK as drug target (Ruderman & Prentki, 2004). AMPK activation might help patients with type 2 diabetes (Winder & Hardie, 1999). This disease shows symptoms, which would be predicted for decreased AMPK activity. In fact, indirect activation of AMPK by AICAR reduced glucose and insulin levels in the blood of test animals. This implies that AMPK activators could find therapeutic use. Inhibitors of AMPK are also of high interest. Tumor cells are known to have high metabolic activity and AMPK seems to contribute to tolerance of tumor cells to metabolic stress (Kato et al., 2002). AMPK inhibition studies with RNAi indeed stopped tumor growth in mice.

The aim of this project was to select AR proteins binding to AMPK by ribosome display. The selected pool of binders would be the starting point for inhibitor and activator screenings. AMPK is a heterotrimer, consisting of one α-, β- and γ-subunit each. An avi tag was genetically fused to the N-terminus α-subunit for *in vivo* biotinylation expression, purification (Neumann et al., 2003) and subsequent immobilisation during selection rounds and ELISA.

Four rounds of standard ribosome display against AMPK were performed and the pool of selected binders was analysed by crude extract ELISA (selection

and screening procedure as described in Binz et al., 2004)). Several binders were obtained (Fig. 1).

DNA of the 10 most promising binders and of the complete pools were handed over to Nadine Straumann in the group of Theo Wallimann for further characterisation and screening.

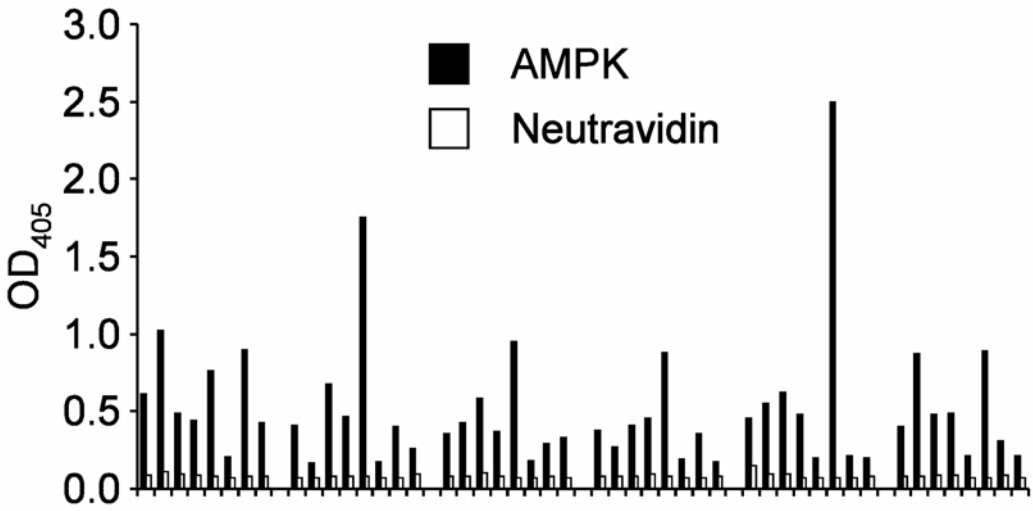


Figure 1. Crude Extract ELISA for identification of NSP binders after four rounds of ribosome display selection. Binding to AMPK (black bar) was compared to binding to an unrelated target protein (grey bar). Of 48 analysed clones 15 gave a strong and specific ELISA signal, while most others (17) still showed significant binding. We concluded that a pool of specific binders was selected, constituting an ideal starting point for the screening of activators and inhibitors.

BACE Inhibitors Selected from Libraries of Designed Ankyrin Repeat Proteins

Hans Kaspar Binz, Oliver Middendorp, Patrick Amstutz, Urs Lüthi, Andreas Plückthun & Alcide Barberis

β-site amyloid precursor protein cleaving enzyme (BACE1) is a membrane bound protease. One distinct cleavage product, a peptide, is responsible for the formation of amyloid plaques and thus is involved in the pathogenesis of Alzheimer's disease I (De Strooper & Konig, 1999; Lin et al., 2000; Vassar et al., 1999). In collaboration with ESBAtech AG (Schlieren, Switzerland) this project aims to generate AR protein based inhibitors. In analogy to the selection of APH inhibitors, a two-step procedure was applied, first selecting a pool of binders by ribosome display, which is subsequently screened for inhibitors in an *in vivo* set-up. We performed four ribosome-display selection rounds against BACE1 (as described in Binz et al. (Binz et al., 2004)) and obtained pools of BACE1-binding AR proteins (see Fig. 1). These pools were further subjected to an *in vivo* screening assay in yeast for the identification of BACE1 inhibitors. Several AR proteins could be isolated that inhibited BACE1 activity *in vivo* in yeast. These inhibitors were then recloned for recombinant expression in *E. coli*. The proteins were produced and purified. Currently, *in vitro* BACE1 inhibition tests are being conducted to demonstrate that indeed enzyme inhibitors have been selected.

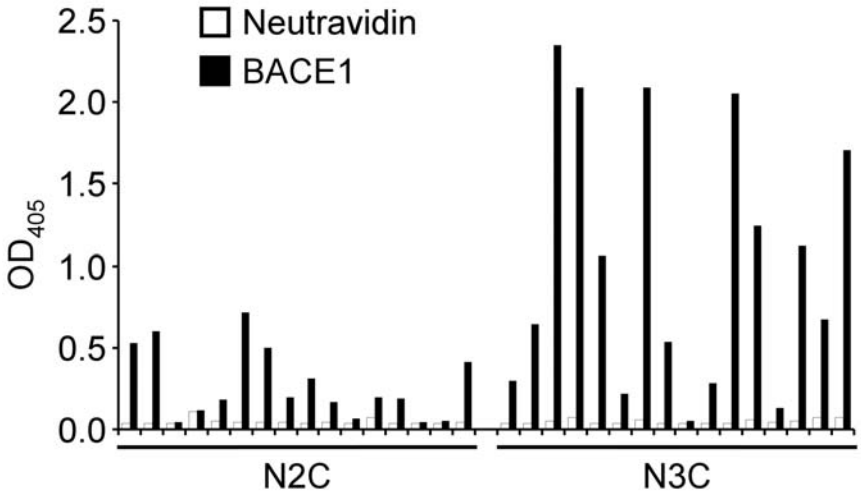


Figure 1. Crude extract ELISA for the identification of BACE1 binding AR proteins of the pools of selected N2C and N3C AR proteins. Of 16 clones from the N2C pool, 11 specifically bound BACE1, one bound unspecifically and 4 gave no signal. Fifteen of 16 clones from one N3C pool specifically bound BACE1 while one clone did not show any signal.

References

Ahola, T. & Kääriäinen, L. (1995). Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl. Acad. Sci. U S A* **92**, 507-511.

Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G. & Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575-582.

Breyton, C., Haase, W., Rapoport, T. A., Kuhlbrandt, W. & Collinson, I. (2002). Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* **418**, 662-665.

De Strooper, B. & Konig, G. (1999). Alzheimer's disease. A firm base for drug development. *Nature* **402**, 471-472.

Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287-294.

Kato, K., Ogura, T., Kishimoto, A., Minegishi, Y., Nakajima, N., Miyazaki, M. & Esumi, H. (2002). Critical roles of AMP-activated protein kinase in constitutive tolerance of cancer cells to nutrient deprivation and tumor formation. *Oncogene* **21**, 6082-6090.

Kemp, B. E., Mitchelhill, K. I., Stapleton, D., Michell, B. J., Chen, Z. P. & Witters, L. A. (1999). Dealing with energy demand: the AMP-activated protein kinase. *Trends Biochem. Sci.* **24**, 22-25.

Kujala, P., Ikaheimonen, A., Ehsani, N., Vihinen, H., Auvinen, P. & Kaariainen, L. (2001). Biogenesis of the Semliki Forest virus RNA replication complex. *J. Virol.* **75**, 3873-3884.

Kujala, P., Rikkonen, M., Ahola, T., Kelve, M., Saarma, M. & Kääriäinen, L. (1997). Monoclonal antibodies specific for Semliki Forest virus replicase protein nsP2. *J. Gen. Virol.* **78 (Pt 2)**, 343-351.

Lin, X., Koelsch, G., Wu, S., Downs, D., Dashti, A. & Tang, J. (2000). Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein. *Proc. Natl. Acad. Sci. U S A* **97**, 1456-1460.

Matlack, K. E., Mothes, W. & Rapoport, T. A. (1998). Protein translocation: tunnel vision. *Cell* **92**, 381-390.

Neumann, D., Woods, A., Carling, D., Wallimann, T. & Schlattner, U. (2003). Mammalian AMP-activated protein kinase: functional, heterotrimeric complexes by co-expression of subunits in Escherichia coli. *Protein Expr. Purif.* **30**, 230-237.

Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. (1995). Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842-846.

Ruderman, N. & Prentki, M. (2004). AMP kinase and malonyl-CoA: targets for therapy of the metabolic syndrome. *Nat Rev Drug Discov* **3**, 340-351.

van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44.

Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., Teplow, D. B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M. A., Biere, A. L., Curran, E., Burgess, T., Louis, J. C., Collins, F., Treanor, J., Rogers, G. & Citron, M. (1999). Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286**, 735-741.

Wang, H. L., O'Rear, J. & Stollar, V. (1996). Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity. *Virology* **217**, 527-531.

Winder, W. W. & Hardie, D. G. (1999). AMP-activated protein kinase, a metabolic master switch: possible roles in type 2 diabetes. *Am. J. Physiol.* **277**, E1-10.

Chapter 8

Discussion, Conclusions and Outlook

Contents

1. Discussion and Conclusions

The results described in this thesis demonstrate that designed AR proteins are a valid alternative to antibodies. The selected AR proteins equal antibodies with respect to their binding characteristics, affinity and specificity, but surpass them in terms of production yield, selection speed and stability, especially under reducing conditions.

The success of designed AR proteins might be explained by their extraordinary biophysical properties (Kohl et al., 2003). A consensus design approach was the foundation for the generation of sequences with these properties (Binz et al., 2003; Forrer et al., 2004). AR proteins based on this sequence are stable, well expressed and show the typical AR fold. These beneficial properties have a positive influence and accelerate experiments at all levels of binder generation, including library quality and selection, and subsequent analysis, i.e. determining specificity, affinity and structure.

Concerning the library quality, the intrinsic stability of the framework sequence is of high importance. A stable framework will tolerate more diversity in the library positions without misfolding. If a library has more functional members, a higher diversity is the consequence. This higher diversity is the key to success for selection experiments. We were indeed able to select not only one binder, but also a whole range of different binders for every target protein. This is especially important if functional properties other than binding are desired. Here, pools of binders may serve as a starting point for functional screens, for example to identify of enzyme inhibitors.

The high stability of the AR proteins also facilitates the actual selection process. Selections for binders from naïve libraries are very challenging. The task is to select the very rare specific binders out of a very large number of non-binders, which might also have "sticky" properties. The better behaved the starting library is or, in other words, the smaller the population of "sticky" proteins in the library, the easier the selection process will be. Misfolded proteins are expected to have exposed hydrophobic patches, sticking to surfaces in an unspecific fashion. In our hands the selection of binders from naïve libraries with ribosome display takes around five to six rounds for antibody libraries (Hanes et

al., 2000), compared to two to three rounds with AR protein libraries. We believe this observation is directly linked to the biophysical characteristics of these proteins.

The stable and rigid backbone of the designed AR proteins might also be beneficial for their binding characteristics. The binding of a selected AR protein to its target protein is a rigid body interaction. Therefore, little entropy is lost upon binding, which might explain the high affinities detected. Furthermore, the specificity of the designed AR proteins might be explained by their rigidity. The selected AR proteins will only bind to the selected surfaces and will not be flexible enough to adapt to the surfaces of other proteins.

The high expression level of AR proteins allows the preparation of large amounts of material in a very simple fashion. Expression in a 100 ml shake flask culture will yield around 10 mg of pure protein, which is enough for detailed analysis of the binding affinity and specificity. This rather small volume allows the handling of around 20 clones in parallel, of which the best can be chosen for further applications. For simple binder screening, the expression volume can be downscaled to less than one ml, allowing the convenient use of 96 well plates.

As AR proteins are devoid of free cysteines, they are intrinsically stable under reducing conditions, as found in the cellular cytoplasm. This does not hold true for antibodies. Proteins not evolved for intracellular stability are often unstable under these conditions and must therefore be especially adapted for such applications (Biocca & Cattaneo, 1995). Our results obtained with AR proteins inside the cell demonstrate their efficacy. This opens new perspectives ranging from target validation to future gene therapy strategies.

Taken together, the rapid and successful generation of all the results described in this thesis reflects in the favorable nature of designed AR proteins

2. Outlook

The results described here may be viewed as a general proof of principle. They provide a foundation for a whole range of future applications in basic research, diagnostics and therapy.

The modular architecture of the repeat proteins, which allows module shuffling, deletion and elongation, is ideal for affinity maturation of binders. This potential will be exploited for Her2 and TNFα binders (Palladino et al., 2003; Yarden & Sliwkowski, 2001). By a conventional error prone PCR approach, picomolar Her2 binders have already been selected (Zahnd et al., unpublished results). The experiments with module alteration are on the way. This might be especially fruitful for TNFα binders, as the epitope can be expanded. Here, not only binders are desired, but inhibitors of receptor binding are the goal (Zahnd et al., unpublished results). An extended epitope will raise the chance of finding a competitor.

The very high specificity of the selected binders along with their intracellular efficacy will allow the study of specific proteins in their natural environment, even if they share high homology to related proteins. The target proteins on hand include kinases, caspases and all proteins with different splice variants. This new level of specificity might allow the specific assignment of activity to individual proteins, not possible to date. Thus drug targets might be validated more accurately. The selected AR proteins might then prove useful as first lead drug compounds or at least accelerate the drug discovery process.

To date, the structures of both AR protein-target complexes which have been under investigation have been solved. This extraordinarily high success rate indicates that AR proteins might be used as crystallization tools. Target proteins, which are highly flexible, might be stabilized in a distinct conformation by an AR binding molecule, allowing crystal growth. Membrane proteins display large hydrophobic areas, compared to rather small hydrophilic ones and are therefore very difficult to crystallize. Antibodies, binding to membrane proteins, enlarge the hydrophilic portion and can propagate crystal growth (Dutzler et al., 2002; Ostermeier et al., 1995). AR proteins should, in principle, be able to do the same.

The applications of AR proteins for diagnostics are at hand. All ELISA based tests could be done with AR proteins. Furthermore, immunohistochemistry and

FACS should also be feasible. Again, the high affinity, specificity and stability of AR proteins should be beneficial. The superior binding characteristics of AR proteins compared to antibodies might even enable more specific and more sensitive diagnostic tests. The high stability of AR proteins, on the other hand, might propagate new test formats, such as protein chips. Such applications might find use in personalized medicine strategies. Experiments in these directions are planned (Binz et al., unpublished results).

Antibody therapeutics is one of the fastest growing product classes of the pharmaceutical industry. If the drug effect is not based on the effector function of the antibody, but on target binding, AR proteins should in principle perform likewise. For tumor targeting the small size of the AR protein might even allow better tumor penetration (Batra et al., 2002). Final answers will only be found in animal experiments. Also the immunogenicity and pharmacokinetics have to be investigated. Experiments in these directions are under way (Stumpp et al., unpublished results).

3. References

Batra, S. K., Jain, M., Wittel, U. A., Chauhan, S. C. & Colcher, D. (2002). Pharmacokinetics and biodistribution of genetically engineered antibodies. *Curr. Opin. Biotechnol.* **13**, 603-608.

Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489-503.

Biocca, S. & Cattaneo, A. (1995). Intracellular immunization: antibody targeting to subcellular compartments. *Trends Cell. Biol.* **5**, 248-252.

Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287-294.

Forrer, P., Binz, H. K., Stumpp, M. T. & Plückthun, A. (2004). Consensus design of repeat proteins. *ChemBioChem* **5**, 183-189.

Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287-1292.

Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. U S A* **100**, 1700-1705.

Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. (1995). Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842-846.

Palladino, M. A., Bahjat, F. R., Theodorakis, E. A. & Moldawer, L. L. (2003). Anti-TNF-alpha therapies: the next generation. *Nat. Rev. Drug Discov.* **2**, 736-746.

Yarden, Y. & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2**, 127-137.

Appendix

Contents

Designed Ankyrin Repeat Proteins: New Tools in Biotechnology

H. Kaspar Binz, Patrick Amstutz, Patrik Forrer, Michael T. Stumpp, Petra Parizek, Christian Zahnd and Andreas Plückthun[1].

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Abstract. We show here that designed ankyrin repeat proteins are powerful binding molecules suitable for a wide range of biotechnological applications. Binders against a great variety of target proteins could be isolated showing specific, high-affinity binding. We demonstrate that selected AR proteins can successfully be used in Western blots, antigen co-purification and in sandwich ELISA experiments. Designed ankyrin repeat proteins are hence a true alternative to antibodies in many aspects.

From Antibodies to Ankyrin Repeat Proteins

Antibodies are renowned for the potential to bind their targets specifically and with high affinity. These characteristics have made antibodies to one of the most useful proteins for biotechnological applications requiring binding molecules. Yet, the difficult recombinant expression of antibodies (often with moderate yields), their complicated composition (several chains, glycosylation) and the dependence of the antibody stability on disulfide bonds has triggered the quest for alternative protein scaffolds that can be used for the generation of binding molecules[1-4]. We have previously introduced designed ankyrin repeat (AR) proteins as alternative binding molecules[5-9]. Using a consensus design strategy[10], we generated combinatorial libraries of designed AR proteins of varying repeat numbers[5]. The analysis of unselected library members revealed that designed AR proteins are unparalleled in terms of recombinant expression yield and thermodynamic stability[5,6]. From the combinatorial AR protein libraries we were able to select binders with high affinity and specificity to maltose binding protein (MBP) or to the mitogen activated protein kinases (MAPKs) JNK2 and p38 (**Table 1**)[8]. The crystal structure of a selected binder in complex with MBP revealed the correct binding mode, i.e. involving the randomized potential interaction residues of the ARs[8]. We further selected inhibitors against an intracellular kinase (aminoglycoside phosphotransferase, APH(3')-IIIa), proving that AR proteins are a valuable alternative to antibody based intrabodies[7].

Here we report on binding molecules selected from combinatorial AR protein libraries and we summarize the binders described earlier (**Table 1**). These results suggest that AR protein libraries can be used for the generation of binding molecules against a great variety of proteins. We demonstrate that designed AR proteins are a real alternative to antibodies for various biotechnological applications. We show for example that selected AR proteins can be used for the co-purification of a target protein from a complex protein mixture. AR proteins can also be used as highly specific primary detection reagents in Western blotting experiments. Analyzing different binders of the same target (APH(3')-IIIa), we found that the binding of the AR proteins is not restricted to one epitope, but different epitopes are targeted by different selected molecules. We took advantage of this fact and designed a sandwich ELISA experiment that allowed the detection of minimal amounts of target protein in a highly concentrated *E. coli* cell extract.

Target diversity

The popularity of antibodies is due to their ability to bind many different target proteins with high affinity and specificity. To be competitive, alternatives to antibodies should perform accordingly. We have shown that libraries of designed AR proteins can be used for the generation of MBP-, JNK2-, p38- and APH(3')-IIIa-binding molecules. Here we show more binders against p38 and JNK2 (**Table 1**). The target-affinities of these binders are – as for the previous binders – in the low nanomolar range (**Table 1**).

Target Co-purification

Immunoprecipitation or target protein co-purification are important for the detection of antigens in cell extracts and for co-crystallization. Considering the high affinity and specificity of AR proteins, it should be possible to efficiently isolate target proteins from cell extracts. We prove this by co-purification of APH(3')-IIIa with a His-tagged APH(3')-IIIa-binding AR protein from an *E. coli* cell extract (**Fig. 1**). In a single IMAC purification step virtually pure protein

[1] Correspondence should be addressed to A.P. (plueckthun@bioc.unizh.ch), Tel: +41 1 635 55 70; Fax: +41 1 635 57 12

Abbreviations: APH(3')-IIIa, Aminoglycoside 3'-phosphotransferase type IIIa; AR, Ankyrin repeat; ELISA, Enzyme-linked immunosorbent assay; IMAC, Immobilized metal affinity chromatography; IPTG Isopropyl-β-D-thiogalactopyranoside; MAPK, Mitogen-activated protein kinase; MBP, Maltose binding protein; Ni-NTA, Nickel-nitrilotriacetic acid; SDS-PAGE, Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Keywords: Ankyrin repeat, Protein design, Protein purification, Sandwich ELISA, Scaffold, Western blot

Table 1. Affinities of AR proteins selected against different targets.

Target	Selected AR protein	Size[a]	K_D [M]	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	Reference
MBP	m3_16	N2C	$1.7 \cdot 10^{-8}$	$6.0 \cdot 10^{5}$	$1.0 \cdot 10^{-2}$	8
	off7	N3C	$4.4 \cdot 10^{-9}$	$4.2 \cdot 10^{5}$	$1.9 \cdot 10^{-3}$	8
	m3_5	N3C	$2.2 \cdot 10^{-8}$	$2.0 \cdot 10^{5}$	$4.4 \cdot 10^{-3}$	8
JNK2	JNK2_2_3	N2C	$2.1 \cdot 10^{-9}$	$9.7 \cdot 10^{5}$	$2.0 \cdot 10^{-3}$	8
	JNK2_2_4	N2C	$5.2 \cdot 10^{-9}$	$8.8 \cdot 10^{5}$	$4.6 \cdot 10^{-3}$	b
p38	p38_2_3	N2C	$3.7 \cdot 10^{-9}$	$9.5 \cdot 10^{5}$	$3.5 \cdot 10^{-3}$	8
	p38_3_8	N3C	$3.7 \cdot 10^{-9}$	$9.8 \cdot 10^{5}$	$3.6 \cdot 10^{-3}$	b
APH(3')-IIIa	3a	N3C	$1.8 \cdot 10^{-9}$	$1.6 \cdot 10^{6}$	$2.7 \cdot 10^{-3}$	7
	3b	N3C	$5.3 \cdot 10^{-10}$	$2.5 \cdot 10^{6}$	$1.1 \cdot 10^{-3}$	7
	#1	N2C	$1.9 \cdot 10^{-8}$	$1.2 \cdot 10^{5}$	$2.2 \cdot 10^{-3}$	7
	#2	N3C	$8.1 \cdot 10^{-9}$	$2.8 \cdot 10^{5}$	$2.3 \cdot 10^{-3.}$	7
	B	N3C	$2.8 \cdot 10^{-8}$	$1.4 \cdot 10^{5}$	$4.1 \cdot 10^{-3}$	7

[a] N2C and N3C represent AR proteins composed of an N-terminal capping repeat, two or three designed AR modules and a C-terminal capping repeat
[b] this study
Materials: The N2C and N3C AR protein libraries as well as the cloning, expression and purification of the target proteins have been described (see ref. [8] and ref. [7]). The ribosome display selection procedure, the screening for binders in the selected AR protein pools by ELISA, quantitative ELISA and kinetic BIAcore measurements of binding AR proteins were done as described[8].

complex could be isolated that contained APH(3')-IIIa and the AR protein in a 1:1 stoichiometry (**Fig. 1**). This stoichiometry was further confirmed by subsequent gel-filtration experiments (data not shown).

Western Blotting

One of the most important biotechnological applications of binding molecules is Western blotting. Antibodies are most frequently used as detection molecules in Western blots. Other detection molecules are not routinely used, unless they recognize a specific tag sequence that was added to a protein of interest on purpose (for example streptavidin/strep-tag)[11]. We show here that designed AR proteins are a true alternative to antibodies for Western blotting (**Fig. 2**). Two APH(3')-IIIa-binding AR proteins were tested for their ability to specifically bind APH(3')-IIIa in a crude *E. coli* cell extract. Both AR proteins detected APH(3')-IIIa specifically (**Fig. 2**). One high-affinity AR protein detected APH(3')-IIIa even at levels where no band for APH(3')-IIIa was observed on the SDS-PAGE (uninduced *E. coli* extract; **Fig. 2**). The lower-affinity AR protein detected APH(3')-IIIa only at high concentrations, where the APH(3')-IIIa-band was also visible on the SDS-PAGE (**Fig. 2**). A Western blot with a control AR protein showed no observable APH(3')-IIIa detection signal (**Fig. 2**).

The finding that AR proteins can be used in Western blotting is not self-evident. Similar to many antibodies, which recognize structural epitopes, an alternative scaffold might also not be suited for Western blotting, since the recognized epitope may no longer be accessible under PAGE or blotting conditions (e.g. discontinuous epitope). Especially for AR proteins, which do not have a groove comparable to antibodies for the interaction with a continuous epitope, it is surprising that Western blotting is feasible. For one APH(3')-IIIa-binding AR protein (#3a, see **Fig. 2**), we were able to determine the crystal structure in complex with APH(3')-IIIa[12]. In the complex, the AR protein is bound to a structural, but rather continuous epitope of APH(3')-IIIa, i.e. mostly a helix close to the kanamycin binding site. This helix is probably also accessible under blotting conditions or it may refold during the blocking procedure. Note that the sensitivity level of the Western blot has not been investigated.

Sandwich ELISA

Sandwich ELISA is a powerful technology to specifically detect a target molecule in a complex protein solution (e.g. *E. coli* cell extract). A sandwich ELISA requires two binding molecules that recognize different epitopes of a target. One binding molecule is immobilized on a support, which is then incubated with the protein mixture including

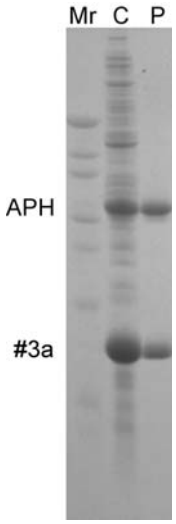


Fig. 1. AR proteins as co-purification agents. The co-purification of APH(3')-IIIa using the selected APH(3')-IIIa-binding AR protein #3a is illustrated by a 15% SDS-PAGE. Mr: molecular marker. C: soluble fraction of a crude *Escherichia coli* (*E. coli*) extract containing a 1:1 mixture of an *E. coli* culture expressing the His-tagged AR protein #3a and a culture expressing untagged APH(3')-IIIa. P: Single step IMAC-purified complex of APH(3')-IIIa and the AR protein #3a from mixture C.

Materials: The AR protein #3a[7] and APH(3')-IIIa were expressed each on a 1.5 l scale as described[5-7]. Before cell lysis, the cell pellets of both cultures were mixed (1:1 ratio). The cell lysis and protein purification has been carried out as described[12].

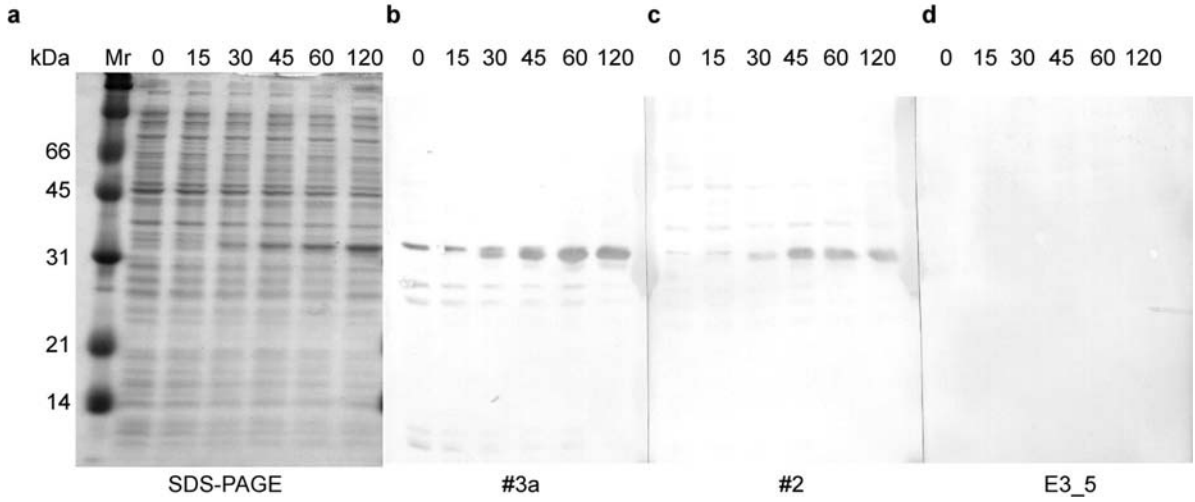


Fig. 2. Application of designed AR proteins in Western blots. (**a**) 15% SDS-PAGE of crude cell extracts of *Escherichia coli* (*E. coli*) expressing APH(3')-IIIa. Samples were taken at different time-points after induction of APH(3')-IIIa expression with IPTG, as indicated above the gel picture (0, 15, 30, 45, 60 and 120 minutes). APH(3')-IIIa appears visibly at approximately 32 kDa after approximately 15 minutes of induction. Mr: molecular weight marker (**b**, **c**, **d**) Western blots of the APH(3')-IIIa samples of (a) with different designed AR proteins. (**b**) Western blot with #3a[7], a N3C AR protein selected to bind and inhibit APH(3')-IIIa. (**c**) Western blot with #2[7], another N3C AR protein selected to bind and inhibit APH(3')-IIIa. (**d**) Western blot with E3_5, an unselected N3C AR protein library member[5].

Materials: The AR proteins, unselected or selected to bind APH(3')-IIIa, as well as their expression and purification have been described[5-7]. APH(3')-IIIa was expressed as described[7] and aliquots were taken at the indicated time-points (see (a)). After 15% SDS-PAGE, the samples were transferred onto an Immobilon-P transfer membrane (Millipore, Billerica, MA, USA) in transfer buffer (12.5 mM Tris base, 96 mM glycine, 10% methanol, pH 8.3) using electroblotting. The membranes were blocked with 0.5% BSA in TBS (50 mM Tris-HCl pH 7.4, 150 mM NaCl) for 180 minutes at room temperature. Between all following incubation steps (all 30 minutes, room temperature), the membranes were extensively washed with TBST (TBS supplemented with 0.05% Tween 20). The membranes were incubated with 100 nM purified AR protein #3a (or #2 or E3_5) in TBS buffer containing 0.5% BSA. For detection of the bound AR proteins, the membranes were then incubated with an αRGS-His IgG (QIAgen) in TBS containing 0.5% BSA. Next, the membranes were incubated with an αmouse-IgG goat IgG-AP fusion (Sigma, St. Louis, MO, USA) in TBS containing 0.5% BSA. Finally, the blots were developed using Nitroblue tetrazoliumchloride (NBT) and 5-bromo-4-chloro-3-indolylphosphate (BCIP) as substrates (100 µl of each NBT and BCIP stocks in 10 ml 100 mM Tris-HCl pH 9.5, 0.5 mM $MgCl_2$; NBT stock: 30 mg/ml in 70% dimethylformamide; BCIP stock: 15 mg/ml in dimethylformamide).

---

the target protein. After extensive washing, the second binding molecule is used for the detection of the target protein bound by the first binding molecule (**Fig. 3**). The high sequence diversity in the pool of APH(3')-IIIa-binding AR proteins (see ref. [7]) indicated that the individual binders recognize different epitopes[7]. Indeed, we could identify AR proteins that bound different epitopes of APH(3')-IIIa (data not shown). One pair was chosen for a quantitative sandwich ELISA analysis of APH(3')-IIIa binding (**Fig. 3**). In the sandwich ELISA, the two AR proteins specifically recognized APH(3')-IIIa in a highly concentrated *E. coli* cell extract at APH(3')-IIIa concentrations far below the detection limit on SDS-PAGE (significant signal at 500 pM; **Fig. 3**). The detection signal was linear with the APH(3')-IIIa concentration in a range from at least 1 nM to 200 nM (**Fig. 3**). The control reactions showed that a signal was only observed, if the two AR proteins and APH(3')-IIIa were present, but not if any of these components was missing. The sensitivity of this sandwich ELISA depends on the affinity of the binders for the target. Hence, the sensitivity of this assay might be increased by using affinity-matured binders or by increasing the avidity of the binders. In this manner designed AR proteins could represent powerful tools for protein-chip applications.

In conclusion, we have further validated designed AR proteins as binding molecules. More binders have been isolated and characterized, and the technical applicability of designed AR proteins was proven to be equivalent to antibodies. Antigen co-purification, Western blotting and sandwich ELISA could all successfully be made using designed AR proteins.

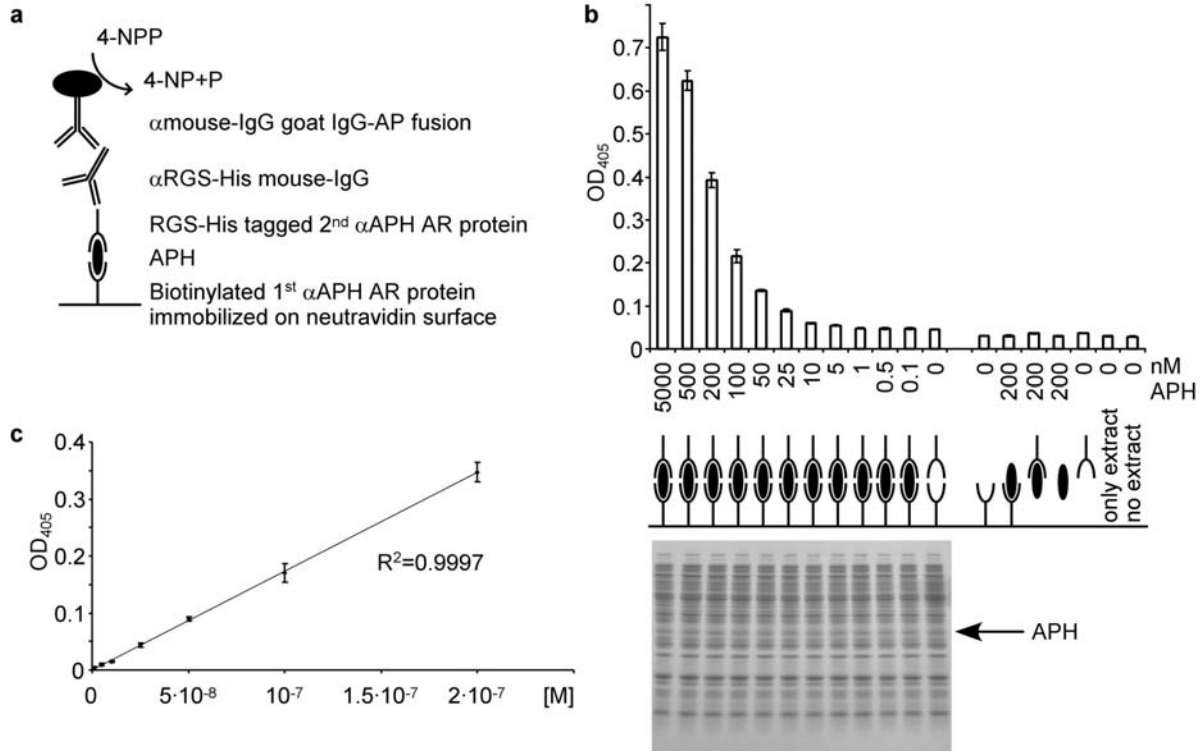


Fig. 3. Sandwich ELISA with two selected AR proteins detecting monomeric APH(3')-IIIa in a crude *E. coli* extract. (**a**) Schematic drawing of the sandwich ELISA and the detection reagents used. Biotinylated AR protein #3b was immobilized on neutravidin-coated plates. A highly concentrated *E. coli* extract spiked with a low concentration of monomeric APH(3')-IIIa was then added. A second AR protein (#B), which binds a different epitope of APH(3')-IIIa than the one of protein #3b, was used to detect the presence of bound monomeric APH(3')-IIIa. The RGS-His-tag of this second AR protein was then detected using an antibody detection system. (**b**) Sandwich ELISA signals obtained with different monomeric APH(3')-IIIa concentrations in the *E. coli* extract (5 μM to 100 pM). The concentrations of monomeric APH(3')-IIIa used are indicated below the signals. Note that in this representation the background signals of the detection antibodies have not been subtracted. Below, the combination of detection molecules used in the sandwich ELISA is depicted describing the setup of the measurement samples (leftmost eleven signals) and the control samples (rightmost eight signals). The 15% SDS-PAGE at the bottom depicts the crude *E. coli* extracts containing the different concentrations of monomeric APH(3')-IIIa used in the sandwich ELISA. (**c**) Linearity of the monomeric APH(3')-IIIa detection signal for the range of 1 nM to 200 nM monomeric APH(3')-IIIa in the sandwich ELISA. For this representation, the background value (no monomeric APH(3')-IIIa in the *E. coli* extract; 12th signal of (a) from the left) was subtracted from the measured values.

Materials: Monomeric APH(3')-IIIa and the RGS-His tagged AR protein #B were produced and purified as described[5-7]. The biotinylated AR protein #3b was prepared as follows: Plasmid pQE30 (QIAgen, Hilden, Germany) harboring #3b[7] was cut with *Bam*HI/*Hin*dIII and the AR protein encoding DNA was ligated into pAT222[8] cut previously with the same enzymes. The resulting vector pAT222_#3b encodes the fusion protein Avi-tag:pD:AR protein:His_6-tag under the control of a T5 promoter. pAT222_#3b and the BirA encoding plasmid pBirAcm (Avidity, Denver, CO, USA) were cotransformed into *E. coli* XL1-Blue (Stratagene, La Jolla, CA, USA), and *in vivo* biotinylated #3b AR protein was produced and purified as described for biotinylated MBP[8]. The *E. coli* cell extract was prepared as follows: *E. coli* BL21-Gold (Stratagene) was transformed with pAT118[13] encoding a truncated form of phage lambda capsid protein D. Seventy five ml culture (LB, 1% glucose, 100 mg/l ampicilin) grown at 37°C overnight were used to inoculate a 1.5 l culture of the same medium. After 7 hours growth at 37°C (without induction of the protein expression), the culture was harvested by centrifugation and the cells were resuspended in 50 ml TBS500 (50 mM Tris-HCl pH 8.0, 500 mM NaCl). After cell lysis with an Emulsiflex C5 (Avestin) followed by additional sonication, the cell extract was centrifuged for 20 min at 50'000 g. The supernatant, spiked with different concentrations of APH(3')-IIIa, was then used for the ELISA. The ELISA was performed as follows: A Maxisorp plate (Nunc, Roskilde, Denmark) was coated with neutravidin and blocked with BSA as described[8]. Saturating amounts of biotinylated #3b AR protein in TBST (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% Tween 20) was added to the required wells (100 μl) for 1 hour at 4°C. Between the following incubation steps (all 1 hour at 4°C), the plate was always washed extensively with ice-cold TBST. The wells were incubated with 100 μl cell extracts containing different concentrations of APH(3')-IIIa. Next, 100 μl (2 μM) RGS-His-tagged #B AR protein were added to the required wells (in TBST containing 0.5% BSA). Following this, 100 μl αRGS-His antibody (QIAgen) in the same buffer were added to the wells according to the manufacturer. Finally, the wells were incubated with 100 μl of the secondary antibody (αmouse-IgG goat IgG-AP fusion; Sigma), again in the same buffer. The ELISA was developed using para-nitrophenylphosphate (4-NPP; Fluka, Buchs, Switzerland) as described[8].

Acknowledgments

HKB was the recipient of a predoctoral fellowship of the Roche Research Foundation. MTS was supported by the „Bundesministerium für Bildung und Forschung“ and the “Fonds der chemischen Industrie”.

References

1. Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M., & Nygren, P.-Å. (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nat. Biotechnol. **15**, 772-777.
2. Nygren, P.-Å., & Uhlén, M. (1997). Scaffolds for engineering novel binding sites in proteins. Curr. Opin. Struct. Biol. **7**, 463-469.
3. Skerra, A. (2000). Engineered protein scaffolds for molecular recognition. J. Mol. Recognit. **13**, 167-187.
4. Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L., et al. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. Chem. Biol. **9**, 933-942.
5. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J. Mol. Biol. **332**, 489-503.
6. Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Plückthun, A., & Grütter, M.G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc. Natl Acad. Sci. USA **100**, 1700-1705.
7. Amstutz, P., Binz, H.K., Parizek, P., Forrer, P., Stumpp, M.T., & Plückthun, A. (2004). Designed ankyrin repeat proteins as intracellular inhibitors. Manuscript.
8. Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grütter, M.G., & Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. Nat. Biotechnol. **22**, 575-582.
9. Binz, H.K., Kohl, A., Plückthun, A., & Grütter, M.G. (2004). 1.9 Å crystal structure of a consensus-designed ankyrin repeat protein. Manuscript.
10. Forrer, P., Stumpp, M.T., Binz, H.K., & Plückthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett. **539**, 2-6.
11. Skerra, A., & Schmidt, T.G. (2000). Use of the Strep-Tag and streptavidin for detection and purification of recombinant proteins. Methods Enzymol. **326**, 271-304.
12. Kohl, A., Amstutz, P., Parizek, P., Binz, H.K., Briand, C., Capitani, G., Forrer, P., Plückthun, A., & Grütter, M.G. (2004). Crystal structure of a complex between a kinase and its designed ankyrin repeat protein inhibitor. Manuscript.
13. Yang, F., Forrer, P., Dauter, Z., Conway, J.F., Cheng, N., Cerritelli, M.E., Steven, A.C., Plückthun, A., & Wlodawer, A. (2000). Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat. Struct. Biol. **7**, 230-237.

J|A|C|S
ARTICLES
Published on Web 07/17/2002

In Vitro Selection for Catalytic Activity with Ribosome Display

Patrick Amstutz,[†] Joelle N. Pelletier,[†,‡] Armin Guggisberg,[§] Lutz Jermutus,[†,] Sandro Cesaro-Tadic,[†] Christian Zahnd,[†] and Andreas Plückthun*[,†]

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and Organisch-Chemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received February 8, 2002

Abstract: We report what is, to our knowledge, the first in vitro selection for catalytic activity based on catalytic turnover by using ribosome display, a method which does not involve living cells at any step. RTEM-β-lactamase was functionally displayed on ribosomes as a complex with its encoding mRNA. We designed and synthesized a mechanism-based inhibitor of β-lactamase, biotinylated ampicillin sulfone, appropriate for selection of catalytic activity of the ribosome-displayed β-lactamase. This derivative of ampicillin inactivated β-lactamase in a specific and irreversible manner. Under appropriate selection conditions, active RTEM-β-lactamase was enriched relative to an inactive point mutant over 100-fold per ribosome display selection cycle. Selection for binding, carried out with β-lactamase inhibitory protein (BLIP), gave results similar to selection with the suicide inhibitor, indicating that ribosome display is similarly efficient in catalytic activity and affinity selections. In the future, the capacity to select directly for enzymatic activity using an entirely in vitro process may allow for a significant increase in the explorable sequence space relative to existing strategies.

Naturally occurring enzymes catalyze a wide variety of chemical reactions and are increasingly used in pharmaceutical, industrial, and environmental applications as a result of their high reactivities and specificities. However, the direct improvement of biocatalysts remains challenging, and the yet more ambitious goal of developing enzymes with new catalytic functions still seems almost elusive. Although our knowledge of structure−function relationships of enzymes has significantly increased, rational protein design is still a difficult task, especially for improved catalysis. The recently developed strategy of directed evolution can be used as a complement to rational design. In directed evolution, a protein function of interest is evolved in the laboratory by mimicking Darwinian evolution in multiple successive rounds of diversification (library generation) with subsequent selection or screening (reviewed in refs 1−3).

Screening, which involves the analysis of single protein variants, can be automated for high-throughput protocols but remains laborious and time-consuming, therefore limiting the size of libraries which can be handled.[4] As opposed to screening, selection methods sample the entire library in a single experimental step. Such experiments require a direct coupling of the phenotype, which is to be selected for, and its encoding genetic information, the genotype.

For the selection of enzymatic activities, three general approaches can be distinguished: methods performed entirely in vivo, those working "partially" in vitro, and those performed completely in vitro.[1] In the in vivo approach, a genetic library encoding enzyme variants is transformed into cells where the variants are expressed and selection takes place. Because selection protocols are generally based on a growth advantage, e.g. complementation of an auxotrophy or resistance to a cytotoxic compound,[5−8] in vivo selection of catalysis is limited to those activities giving rise to a growth advantage. Moreover, microbial genomes have evolved to deal with environmental selection pressure. The expression host can therefore be surprisingly "creative" in escaping selection pressure, such as by higher expression of a poor catalyst or the use of alternative metabolic pathways, completely by-passing the enzymatic activity that is the actual target of selection.

Partially in vitro methods can offer an alternative to in vivo methods. In these methods, the library is also introduced into cells, resulting in display of the protein of interest, usually on the surface of phage,[9] bacteria, or yeast.[10] Selection then occurs

* Corresponding author. Tel. (+41-1) 635 55 70. Fax: (+41-1) 635 57 12. E-mail: plueckthun@biocfebs.unizh.ch.
[†] Biochemisches Institut, Universität Zürich.
[§] Organisch-Chemisches Institut der Universität Zürich.
[‡] Present address: Département de Chimie, Université de Montréal, C. P. 6128, Succursale Centre-ville, Montréal, PQ, Canada.
Present address: Cambridge Antibody Technology, The Science Park, Melbourn, Cambridgeshire SG8 6JJ, U.K.

(1) Griffiths, A. D.; Tawfik, D. S. *Curr. Opin. Biotechnol.* **2000**, *11*, 338−353.
(2) Olsen, M.; Iverson, B.; Georgiou, G. *Curr. Opin. Biotechnol.* **2000**, *11*, 331−337.
(3) Soumillion, P.; Fastrez, F. *Curr. Opin. Biotechnol.* **2001**, *12*, 387−394.
(4) Joo, H.; Lin, Z.; Arnold, F. H. *Nature* **1999**, *399*, 670−673.
(5) Orencia, M. C.; Yoon, J. S.; Ness, J. E.; Stemmer, W. P.; Stevens, R. C. *Nat. Struct. Biol.* **2001**, *8*, 238−242.
(6) Altamirano, M. M.; Blackburn, J. M.; Aguayo, C.; Fersht, A. R. *Nature* **2000**, *403*, 617−622. Retraction, *Nature* **2002**, *417*, 468.
(7) Crameri, A.; Raillard, S. A.; Bermudez, E.; Stemmer, W. P. *Nature* **1998**, *391*, 288−291.
(8) MacBeath, G.; Kast, P.; Hilvert, D. *Science* **1998**, *279*, 1958−1961.
(9) Dunn, I. S. *Curr. Opin. Biotechnol.* **1996**, *7*, 547−553.
(10) Mendelsohn, A. R.; Brent, R. *Science* **1999**, *284*, 1948−1950.

10.1021/ja025870q CCC: $22.00

in vitro, that is, outside the cell, allowing fine-tuning of the selection pressure and selection conditions.

Both entirely in vivo or partially in vitro techniques require a transformation step, limiting the applicable library size to cellular transformation efficiencies. Typically, transformation efficiencies of 10^7-10^8 cells/μg DNA for yeast and 10^9-10^{10} cells/μg DNA for *Escherichia coli* (*E. coli*)[11,12] are achievable. A considerable amount of work arises from the need of ligating and transforming such a library after each round of randomization. Furthermore, cytotoxic proteins cannot be displayed at all. Technologies working completely in vitro can overcome these limitations, because no living cells are involved at any step.

Two approaches that are carried out completely in vitro can be distinguished. In one, a compartmentalization of individual variants is achieved in water-in-oil emulsions.[13] Since the in situ detection of fluorescent products and direct optical sorting of such droplets, harboring the catalytic proteins, have not yet been reported, the physical link between genotype and phenotype is still required, and the emulsions have to be broken up before an affinity selection or sorting can be performed. The other approach makes use of the concomitant presence of mRNA and nascent protein at the ribosome during an in vitro translation for coupling of genotype and phenotype.[14] Here, the most prominent techniques are ribosome display;[15,16] mRNA display, also termed "in vitro virus"[17] or "mRNA−protein fusion";[18] and "ribosome-inactivation display system".[19] Below, we will describe the first application of ribosome display to the selection of catalytic activity in an effort to increase the library sizes accessible beyond those in previously published accounts with in vivo or partially in vitro methods.

To select catalysts in vitro, catalysis must be correlated or coupled to binding. Although reversible binding of enzymes to transition state analogues has been used frequently in the selection of novel catalysts,[20,21] a more direct selection for enzymatic turnover can be achieved by the use of mechanism-based (or "suicide") inhibitors.[22] These compounds are substrate analogues that, upon turnover, are converted into a reactive species that binds with the enzyme in a covalent manner, thus causing irreversible inhibition. Mechanism-based inhibitors can thus be used for labeling active enzyme molecules. The selection of β-lactamase displayed on phage with a biotinylated mechanism-based inhibitor has been reported.[23−26] In these reports, active enzyme was enriched relative to less active mutants and relative to penicillin binding protein. The elution of covalently trapped clones from the matrix to which they had been linked is generally achieved by cleaving within a linker region, either chemically between the suicide inhibitor and the affinity tag or enzymatically between the displayed protein and the phage.

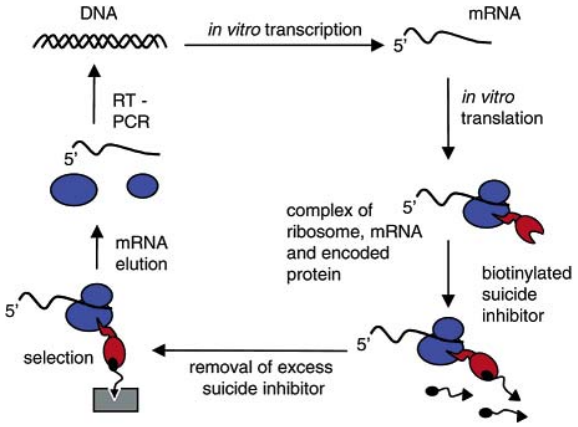


Figure 1. Principle of ribosome display selection for catalytic activity. DNA encoding the protein of interest is transcribed and translated in vitro. Because the mRNA carries no stop codon and translation is stopped with high Mg^{2+} concentrations, stable ternary complexes of mRNA (black), ribosome (blue) and tethered nascent protein (red) are formed. These can be used directly for selection with a biotinylated suicide inhibitor. Excess inhibitor is removed by gel filtration, and the labeled complexes are captured with avidin coated magnetic beads. After washing to remove any untrapped ribosomal complexes, the selected complexes are destroyed to elute the mRNA. Finally, reverse transcription (RT) and PCR are used to amplify the genetic information of the selected clones.

Even though entirely in vitro display techniques have great potential for selection of catalytic activity, this has not previously been reported. We believe that ribosome display is particularly well suited to applications based on mechanism-based inhibition, because ribosome display offers an elegant solution for elution of the selected genetic information. Because genotype and phenotype are coupled in a noncovalent complex at the ribosome (Figure 1), dissociation of these complexes and therefore elution of the genetic information is easily achieved.

In the present study, we use ribosome display for selection of catalytic activity with a suicide inhibitor. We have chosen β-lactamase as a model enzyme, and we displayed it on the ribosome in complex with its encoding mRNA. We have devised a synthetic scheme for the preparation of a biotinylated mechanism-based inhibitor of β-lactamase. We then applied the inhibitor to the selective enrichment of active, displayed enzyme. We show that selection of enzymes is possible entirely in vitro, combining the advantages of both the in vitro technology of ribosome display and direct selection for catalytic turnover with a suicide substrate to select for protein catalysts.

Results and Discussion

In this study we assessed the potential of ribosome display for enzyme selection based on catalytic activity and on binding specificity. In such an in vitro selection technology, the theoretical library size would be limited only by the number of active ribosomes in the reaction. As a model system, we displayed RTEM β-lactamase in a ribosomal complex with its encoding mRNA. We describe a rapid synthesis of a biotinylated mechanism-based β-lactamase inhibitor for activity selection.

(11) Inoue, H.; Nojima, H.; Okayama, H. *Gene* **1990**, *96*, 23−28.
(12) Sidhu, S. S.; Lowman, H. B.; Cunningham, B. C.; Wells, J. A. *Methods Enzymol.* **2000**, *328*, 333−363.
(13) Tawfik, D. S.; Griffiths, A. D. *Nat. Biotechnol.* **1998**, *16*, 652−656.
(14) Amstutz, P.; Forrer, P.; Zahnd, C.; Plückthun, A. *Curr. Opin. Biotechnol.* **2001**, *12*, 400−405.
(15) Hanes, J.; Plückthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4937−4942.
(16) He, M.; Taussig, M. J. *Nucleic Acids Res.* **1997**, *25*, 5132−5134.
(17) Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. *FEBS Lett.* **1997**, *414*, 405−408.
(18) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297−12302.
(19) Zhou, J. M.; Fujita, S.; Warashina, M.; Baba, T.; Taira, K. *J. Am. Chem. Soc.* **2002**, *124*, 538−543.
(20) Schultz, P. G.; Lerner, R. A. *Science* **1995**, *269*, 1835−1842.
(21) Xu, J.; Deng, Q.; Chen, J.; Houk, K. N.; Bartek, J.; Hilvert, D.; Wilson, I. A. *Science* **1999**, *286*, 2345−2348.
(22) Hilvert, D. *Annu. Rev. Biochem.* **2000**, *69*, 751−793.
(23) Soumillion, P.; Jespers, L.; Bouchet, M.; Marchand-Brynaert, J.; Winter, G.; Fastrez, J. *J. Mol. Biol.* **1994**, *237*, 415−422.
(24) Vanwetswinkel, S.; Marchand-Brynaert, J.; Fastrez, J. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 239.
(25) Vanwetswinkel, S.; Avalle, B.; Fastrez, J. *J. Mol. Biol.* **2000**, *295*, 527−540.
(26) Avalle, B.; Vanwetswinkel, S.; Fastrez, J. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 479−484.

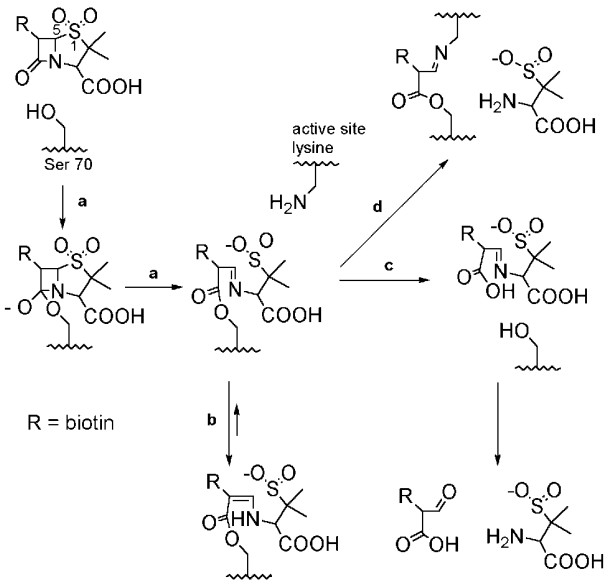


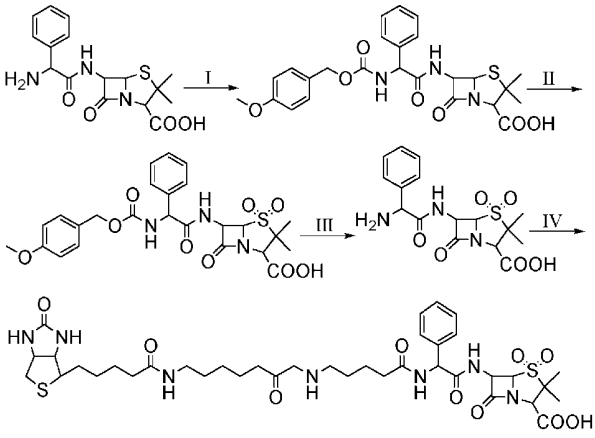
Figure 2. Proposed interaction of ampicillin sulfone with the active site serine β-lactamase.[27,28] The active site serine attacks the carbonyl group of the lactam ring, and before a water molecule can attack, the sulfone in position 1 acts as a leaving group from carbon 5 and evokes the formation of an acyl-enzyme intermediate (a). This acyl-enzyme intermediate is in an equilibrium with a more stable tautomer (b), resulting in transient inhibition. A normal deacylation, triggered by water, can also occur, and no inhibition is seen (c). Finally, transamination with a lysine in the active site can occur (d), which results in an irreversibly inactivated enzyme, bound covalently to the inhibitor and the biotin moiety.

We demonstrate that the enzyme folds on the ribosome to its correct three-dimensional structure and is active there. Furthermore, we demonstrate selection based on activity as well as selection for affinity with this system.

Rapid Synthesis of a Bifunctional Mechanism-Based Inhibitor: Biotinylated Ampicillin Sulfone. Penicillin analogues with an electron withdrawing group in position 1 (Figure 2) are known to lead to opening of the thiazolidine ring after nucleophilic attack of the β-lactam ring by the active-site serine of type-A β-lactamase, forming an acyl-enzyme intermediate.[27,28] The attack of a second active-site nucleophilic side chain on this reactive intermediate results in a covalently inhibited enzyme (Figure 2). Clavulanic acid and penicillanic acid sulfone are well-characterized mechanism-based inhibitors of type-A β-lactamase that follow this mechanism.[29,30] By tethering biotin to such an inhibitor, an affinity handle is generated that links biotin exclusively to active enzyme molecules as a consequence of enzymatic turnover. The resulting irreversible trapping allows for the possibility to select directly for catalytic activity.[23]

We developed a four-step synthesis for biotinylated ampicillin sulfone, a β-lactamase suicide inhibitor, with 64% overall yield. The synthesis of ampicillin sulfone involved only protection of the amino group of ampicillin, oxidation of the sulfur in the thiazolidine ring to the sulfone, followed by deprotection (Scheme 1). Since the compound is highly sensitive to hydrolysis at the lactam ring, the choice of the protecting group proved to be of utmost importance. Only protection with 4-methoxybenzyloxycarbonyloxyimino-2-phenyl-acetonitrile (MOZ-ON), but not with benzylchloroformate or paranitrobenzylchloroformate, yielded intact ampicillin sulfone upon deprotecting. In a fourth step, the biotin moiety was added to yield the biotinylated suicide inhibitor (Scheme 1). Our synthesis is simpler than that of a similar penicillanic acid sulfone, because we do not need a disulfide bond in the linker region between the penicillin and the biotin moiety.[31] This disulfide bridge was required in the context of phage display selection of β-lactamase to release the covalently trapped phages by cleaving with DTT. In contrast, ribosome display does not require elution of the covalently trapped and displayed protein, because the genetic information coding for the displayed protein can easily be released by chelating the Mg^{2+} ions so as to disrupt the ribosomal complex.

Scheme 1. Synthesis of Biotinylated Ampicillin Sulfone, a Mechanism-Based Inhibitor for Selection of Active Site Serine β-Lactamases.[a]



[a] (I) H_2O, CH_2Cl_2 (4:1); $NaHCO_3$; MOZ-ON. (II) H_2O, $KMnO_4$. (III) CH_2Cl_2, TFA (10:1). (IV) H_2O, EZ-Link Sulfo-NHS-LC-LC-Biotin (Pierce)

Mechanism-Based Enzyme Inhibition. The ability of ampicillin sulfone to irreversibly inhibit β-lactamase was tested by steady-state kinetic assays. The inhibition pattern of the RTEM β-lactamase from *E. coli*, obtained directly from in vitro translation, was compared to that of the β-lactamase from *Enterobacter cloacae*. Both enzymes behaved identically. The enzyme inhibition profile as a function of time showed a dramatic fall of activity within the first 20 min, followed by a slower activity decay (Figure 3). These biphasic inhibition kinetics are typical for mechanism-based inhibitors of β-lactamases, such as clavulanic acid, 6-(methoxymethylene)penicillanic acid and 6-β-((carboxy)methylsulfonamido)penicillanic acid sulfone.[27,32] The kinetics have been interpreted as being due to a first inhibition phase where the enzyme is transiently inhibited, while in a slower second phase, the compound inhibits the enzyme irreversibly (Figure 2). Since not every turnover results in enzyme deactivation (Figure 2), the inhibitor-to-enzyme ratio needed for complete inhibition is a measure for inhibitor potency. From various inhibition experiments, we determined that this ratio is 8×10^4 for ampicillin sulfone, which is about an order of magnitude higher than that for the previously characterized β-lactamase suicide inhibitor sulbactam.[28,33] This reduced inhibition efficiency is most likely due to the fact that

(27) Fisher, J.; Charnas, R. L.; Knowles, J. R. *Biochemistry* **1978**, *17*, 2180−2184.
(28) Fisher, J.; Charnas, R. L.; Bradley, S. M.; Knowles, J. R. *Biochemistry* **1981**, *20*, 2726−2731.
(29) Brenner, D. G.; Knowles, J. R. *Biochemistry* **1984**, *23*, 5833−5839.
(30) Charnas, R. L.; Knowles, J. R. *Biochemistry* **1981**, *20*, 3214−3219.
(31) Vanwetswinkel, S.; Fastrez, J.; Marchand-Brynaert, J. *J. Antibiot.* **1994**, *47*, 1041−1051.
(32) Brenner, D. G.; Knowles, J. R. *Biochemistry* **1981**, *20*, 3680−3687.
(33) Imtiaz, U.; Billings, E. M.; Knox, J. R.; Mobashery, S. *Biochemistry* **1994**, *33*, 5728−5738.

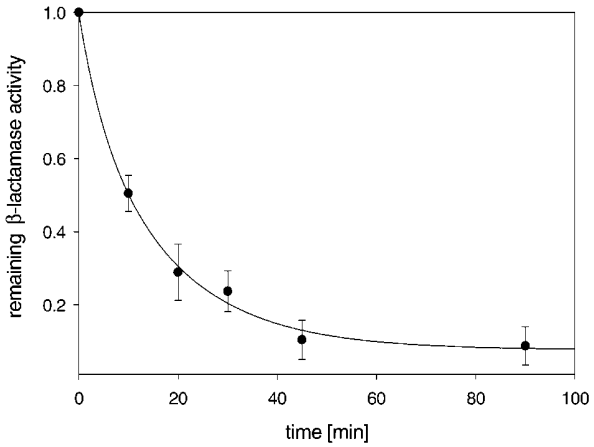


Figure 3. Time course of RTEM-β-lactamase inhibition by ampicillin sulfone. β-lactamase (10^{-9}−10^{-10} M, from in vitro translation) was incubated with ampicillin sulfone (1 mM) for different time periods at room temperature, and the remaining activity was measured in nitrocefin assays. The data, given as activity relative to that at time = 0, were fitted as double exponential decay. Similar inhibition curves have been reported for other β-lactamase suicide inhibitors.[27,31,34] The data for this plot were obtained from three independent experiments.

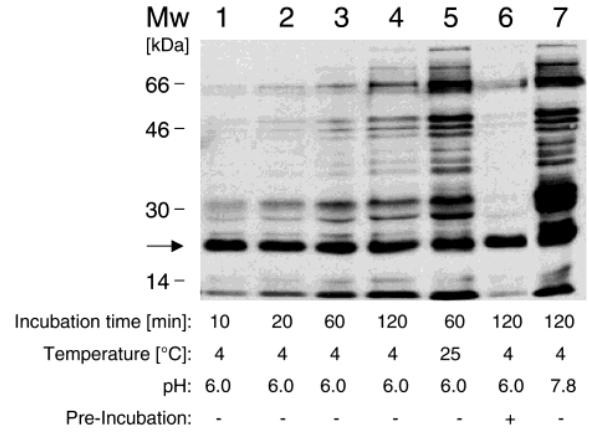


Figure 4. Analysis of nonspecific labeling of proteins by ampicillin sulfone. A protein mix, here *E. coli* S30 extract, was incubated with biotinylated ampicillin sulfone (0.1 mM). The proteins were separated from excess inhibitor by gel filtration, followed by SDS−PAGE and Western blotting for detection of biotin with avidin−alkaline phosphatase conjugate. The arrow indicates the signal obtained from pyruvate carboxylase, a biotinylated *E. coli* protein in the extract, used as internal standard. Lanes 1−5 and 7 show different labeling conditions and incubation times. In lane 6, the protein mix was preincubated with 1 mM ampicillin sulfone at room temperature for 1 h. The nonspecific labeling is most likely caused by the attack of nucleophilic surface residues on the lactam ring and has been reported also for ampicillin.[35]

ampicillin sulfone is derived from ampicillin, a very good substrate of β-lactamase that is rapidly hydrolyzed.[28] The covalent nature of the inhibition was confirmed by dialysis of the inhibited enzyme. Even upon extensive dialysis, no recovery of activity could be detected, whereas the noninhibited enzyme remained fully active (data not shown). The biotinylated form of the inhibitor gave results comparable to ampicillin sulfone itself in all assays (data not shown). These results confirm that the biotinylated form of ampicillin sulfone acts as a bona fide mechanism-based inhibitor of β-lactamase and could be used for selection of β-lactamase activity.

Prevention of Nonspecific Protein Labeling by Ampicillin Sulfone. Besides its reactivity, we also investigated the specificity of the suicide inhibitor. Although hydrolysis of the lactam ring, leading to formation of the reactive species, will be greatly accelerated at the active site of the enzyme, the labile nature of the lactam ring may allow for labeling outside of the active site. In addition, ampicillin is known to react nonspecifically with ϵ-amino groups of proteins in a covalent manner, a cause for ampicillin allergies.[35] Nonspecific labeling could lead to significant background in selection rounds by the suicide inhibitor reacting with inactive enzyme displayed on the ribosome or with ribosomal proteins. We thus investigated possible nonspecific binding in order to minimize it. Aliquots of *E. coli* S30 extract were incubated with biotinylated ampicillin sulfone under various conditions and for different incubation times. The proteins were then separated from excess suicide inhibitor by gel filtration. The resulting samples were analyzed by Western blotting. The results showed significant labeling of proteins other than β-lactamase (Figure 4). Nonspecific labeling also occurred when incubating the biotinylated suicide inhibitor with purified control proteins, such as citrate synthase, GFP (green fluorescent protein), anti-GCN4-single-chain Fv, and M13 helper phage (data not shown). The signal intensity of nonspecific labeling increased with time (lanes 1−4), temperature (lane 5), and higher pH (lane 7) and could be competitively inhibited by preincubation with ampicillin sulfone (lane 6) (Figure 4). These observations are in agreement with the published results on β-lactamase selections with phage display using similar mechanism-based inhibitors.[23−26] In all the phage display experiments, the labeling time was short (<30 min), the pH was kept below 6, and there was at least 1% BSA present, presumably to reduce nonspecific labeling. Under those conditions, the authors reported the selection of highly active β-lactamase over less active mutants or penicillin binding protein, which binds but does not hydrolyze the β-lactam ring.

The broadly based nonspecific labeling of proteins that are not linked to their genotype, as illustrated above, is not likely to be detrimental to the selection of displayed enzyme. However, it suggests that nonspecific labeling of inactive displayed enzyme and of ribosomal proteins may also occur under certain conditions. To favor specific labeling in our in vitro system, the selection for enzymatic activity of β-lactamase was carried out at pH 6.0 and at 4 C for short times (10 min to 1 h) (10 min in Figure 4, lane 1). These conditions are also compatible with those required for ribosome display.

Ribosome-Displayed β-Lactamase is Active. In ribosome display, selection is performed with a ternary complex of mRNA, ribosome and displayed protein (Figure 1). For this strategy to be successful, the protein must fold to its native three-dimensional structure before the polypeptide is released from the ribosome. Furthermore, the ribosome-bound protein must be active. To allow the displayed protein of interest to exit the ribosomal peptide channel and fold into its active conformation, the encoding gene was fused to a 171-amino acid-long C-terminal fusion partner TolA of *E. coli* serving as a tether. To demonstrate functionality of β-lactamase in the ribosomal complex, the complexes were separated by gel filtration from enzyme released during translation, under conditions securing complex stability (see Experimental Procedures). Usually such protein-ribosome complexes are purified by sucrose gradient

(34) Brenner, D. G.; Knowles, J. R. *Biochemistry* **1984**, *23*, 5839−5846.
(35) Schneider, C. H.; De Weck, A. L. *Nature* **1965**, *208*, 57−59.

Table 1. Activity of Ribosome-Displayed β-Lactamase

	sum of free and ribosome-displayed β-lactamase[a]	ribosome-bound β-lactamase[b]	ribosome-displayed β-lactamase after release[c]	background activity in complex fraction[d]
rel enzyme activity[e]	100 ± 4.8	50 ± 5.1	54 ± 7.0	17 ± 5.8

[a] Total activity measured after in vitro translation. [b] After gel filtration (activity of ribosomal complex fraction). [c] Complexes destroyed by RNAse A treatment after gel filtration (activity of ribosomal complex fraction). [d] Complexes destroyed by RNAse A treatment prior to gel filtration (activity of ribosomal complex fraction). [e] Data from three independent experiment measured in triplicates.

centrifugation.[36−38] The use of small gel filtration columns is a rapid and gentle alternative and allows the parallel processing of a large number of samples. Thus, we were able to quantify the amount of active β-lactamase in ribosomal complexes (Table 1). An aliquot of this fraction was treated with RNase A, destroying the ribosomal complex, thus releasing the enzyme. No significant increase in activity was measured, indicating that ribosome-bound β-lactamase is fully active, and no additional active enzyme molecules are obtained by releasing them from the ribosome (Table 1). When the RNase A treatment was carried out prior to the gel filtration procedure, however, some background activity was detected in the complex fraction (Table 1), either from insufficient complex destruction or incomplete separation by gel filtration. These results confirm that the tethered β-lactamase can fold to its correct three-dimensional structure and can catalyze substrate turnover while still bound to the ribosome. Comparing the total activity after in vitro translation of the ribosome display construct prior to gel filtration to the amount of ribosome-displayed activity, we estimated the percentage of β-lactamase in ribosomal complexes to be ~50% of total β-lactamase produced. We presume that the other half is released from the ribosome by hydrolysis of RNA or proteolysis of the tether. The possibility that a fraction of the produced protein is bound to the ribosome in an inactive state and does not refold upon complex destruction remains, but that seems unlikely, because β-lactamase is known to fold efficiently.[39] Although co-translational folding and activity on the ribosome have been demonstrated for firefly luciferase in different experiments,[36−38] we show here that β-lactamase also folds and acquires activity on the ribosome, thereby fulfilling the major requirements for ribosome display selections.

Affinity Selection and Selection for Activity Are Equally Efficient. We extended our analysis of β-lactamase functionality in the ribosome display format by performing selection rounds based on affinity and catalytic activity. The β-lactamase inhibitory protein (BLIP) binds near the active site of β-lactamase with high affinity ($K_d = 0.6$ nM)[40] and had been used in affinity selection with phage display.[41,42] In contrast to His-tagged BLIP, gpHDBLIP, a C-terminal fusion of BLIP to the His-tagged protein D (gpHD),[43] could be expressed at high levels in inclusion bodies and could be refolded efficiently. The fusion partner, gpHD, not only improved the expression of BLIP very significantly but also provides a His tag, which we used for purification of the fusion protein and its immobilization for affinity selection. We performed selection rounds with active ribosome-displayed β-lactamase on either immobilized gpHDBLIP or the immobilized biotinylated suicide inhibitor. The experiments were carried out as described[44] with some modifications to adapt the system for activity selection, notably lower pH (pH 6.0) and the presence of 1% BSA, for reasons discussed above. After in vitro translation and incubation on ice, aliquots of ribosomal complexes displaying β-lactamase were incubated for 1 h with gpHDBLIP or biotinylated ampicillin sulfone. Excess ampicillin sulfone or unbound gpHDBLIP was removed by gel filtration, and the complexes selected with the biotinylated suicide inhibitor were captured with avidin-agarose, whereas the ones binding to gpHDBLIP were immobilized via a tetra-His antibody which was captured with protein G-PLUS agarose. Extensive washing removed unbound complexes. The mRNA was eluted, and after RT and PCR, the yields were quantified according to band intensities on an agarose gel (Figure 5). We observed identical results with affinity- and activity-based selections, indicating that the ribosome-displayed β-lactamase is active in terms of both binding and enzymatic activity and that selection for activity and affinity in this system are equally efficient.

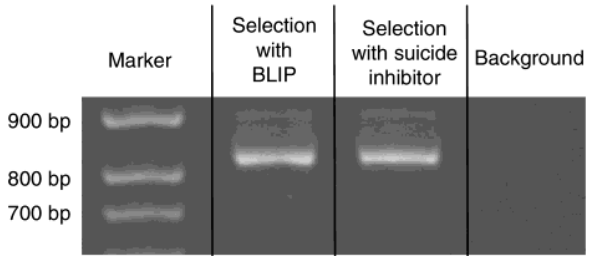


Figure 5. β-Lactamase selection for activity and affinity. The DNA yield after one round of ribosome display by affinity with gpHDBLIP (BLIP), by biotinylated ampicillin sulfone (suicide inhibitor) or with beads alone (background) were compared. Only active β-lactamase was used, resulting in PCR amplification of a band at 834 bp. Lane M indicates DNA molecular weight marker. Background binding to the beads is very weak, and the efficiency of selection for affinity and activity are comparable.

Selection of Active β-Lactamase over an Inactive Mutant Completely In Vitro. To assess the efficiency and specificity of selection for activity with ribosome display, the enrichment of active β-lactamase over an inactive point mutant was tested. The serine residue at position 70 of the active enzyme, whose hydroxyl group carries out the nucleophilic attack of the lactam ring (Figure 2), was replaced by alanine, yielding an inactive mutant enzyme with otherwise unimpaired structure.[45−48] Ri-

(36) Kolb, V. A.; Makeyev, E. V.; Spirin, A. S. *J. Biol. Chem.* **2000**, *275*, 16597−16601.
(37) Makeyev, E. V.; Kolb, V. A.; Spirin, A. S. *FEBS Lett.* **1996**, *378*, 166−170.
(38) Frydman, J.; Erdjument-Bromage, H.; Tempst, P.; Hartl, F. U. *Nat. Struct. Biol.* **1999**, *6*, 697−705.
(39) Vanhove, M.; Guillaume, G.; Ledent, P.; Richards, J. H.; Pain, R. H.; Frere, J. M. *Biochem. J.* **1997**, *321*, 413−417.
(40) Albeck, S.; Schreiber, G. *Biochemistry* **1999**, *38*, 11−21.
(41) Huang, W.; Petrosino, J.; Palzkill, T. *Antimicrob. Agents Chemother.* **1998**, *42*, 2893−2897.
(42) Huang, W.; Zhang, Z.; Palzkill, T. *J. Biol. Chem.* **2000**, *275*, 14964−14968.
(43) Forrer, P.; Jaussi, R. *Gene* **1998**, *224*, 45−52.
(44) Hanes, J.; Jermutus, L.; Plückthun, A. *Methods Enzymol.* **2000**, *328*, 404−430.
(45) Sigal, I. S.; DeGrado, W. F.; Thomas, B. J.; Petteway, S. R., Jr. *J. Biol. Chem.* **1984**, *259*, 5327−5332.
(46) Dalbadie-McFarland, G.; Neitzel, J. J.; Richards, J. H. *Biochemistry* **1986**, *25*, 332−338.

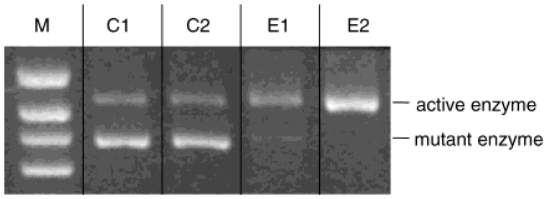


Figure 6. Restriction analysis for detection of enrichment of active β-lactamase with ribosome display. The results of two independent experiments are shown, where E1 represents a usual result and E2, the most successful result obtained. In each experiment, an *Msc*I digest of the DNA pool was carried out after one round of enrichment for enzymatic activity of β-lactamase with ribosome display**.** The round was performed with a ratio of inactive mutant to active β-lactamase of 10:1. After RT, the β-lactamase gene was amplified (834 bp) by PCR and digested with *Msc*I, resulting in cleavage of only the DNA encoding the inactive mutant (662 bp + 172 bp). Lane M indicates DNA molecular weight marker. Lane C1 shows the input mRNA directly reverse-transcribed and amplified by PCR without undergoing a display cycle. C2 shows the background of a standard enrichment round without biotinylated ampicillin sulfone present. E1 and E2 represent two independent enrichment rounds performed with biotinylated ampicillin sulfone. The enrichment factors were estimated to be 110 for E1 and 1400 for E2.

bosome display was performed with RNA encoding active and mutant β-lactamase mixed in a ratio of 1:10. After translation, biotinylated ampicillin sulfone was added, and after 30 min, excess suicide inhibitor was removed by gel filtration. The biotin-labeled complexes were captured with streptavidin magnetic beads, and the mRNA was eluted and amplified by RT and PCR. Five completely independent selection experiments were performed. Enrichment of the active enzyme was quantified by two independent methods: first, an *Msc*I restriction site had been introduced along with the S70A mutation, allowing restriction fragment analysis to distinguish the PCR products encoding active and inactive enzyme. Figure 6 illustrates a representative enrichment (lane E1, factor of 110) as well as the best enrichment obtained (lane E2, factor of 1400). To independently determine the magnitude of the enrichment factor, gene pools after one round of ribosome display were sequenced for the 2 experiments illustrated in Figure 6, and the signal intensities of the respective bases were analyzed, giving enrichment factors of 110 and 250 (data not shown). The fact that the enrichment factor varied within 1 order of magnitude in different selection experiments is most likely due to the small amount of inhibitor used, giving a low yield difficult to quantify, and the inhibitor concentration was not increased to prevent unspecific labeling.

In summary, all experiments clearly showed enrichment of active enzyme over inactive mutant. We concluded that after only one round of mechanism-based selection with ribosome display, an enrichment >100-fold was achieved. To demonstrate that the enrichment was, indeed, due to specific labeling of active β-lactamases, several control experiments were performed. To ascertain that enrichment was not an artifact of RT and PCR, input mRNA was directly applied to this procedure, showing no enrichment of the active species (Figure 6, lane C1). To rule out mRNA stability during translation as enrichment source, a second control was performed. A normal ribosome display round was performed using the same ratio of input mRNA but in the absence of biotinylated ampicillin sulfone. Here, no enrichment of active enzyme was detected (Figure 6, lane C2). The possibility of different translation efficiencies of mutant and active construct was ruled out by performing radioactive translations and comparing protein yields after SDS−PAGE and autoradiography. The band intensities were identical for both constructs (data not shown). It follows that active β-lactamase is efficiently enriched with ribosome display in vitro due to its catalytic activity.

Conclusions

In conclusion, we have shown for the first time that ribosome display may be used for selection of enzymes based on their catalytic activity. It follows that this strategy is ideally suited to performing directed evolution of enzymes. This novel potential of ribosome display was demonstrated by selecting β-lactamase based on its enzymatic activity. We show that the enzyme β-lactamase folds correctly on the ribosome and is active in the ternary complex, thus fulfilling the major requirements for ribosome display. To select for enzymatic activity of β-lactamase with ribosome display, we synthesized ampicillin sulfone, a mechanism-based inhibitor of β-lactamase, which can be derivatized with biotin, yielding a bifunctional activity label. This label was applied to ribosome display experiments, in which active β-lactamase could be enriched over an inactive mutant >100-fold/selection round. The efficiency of selection for activity was comparable to that of affinity selection, which was carried out with immobilized BLIP.

Taken together, we demonstrate that ribosome display, a technique that combines the advantages of very rapid selection rounds with the use of large libraries, known to be capable of evolving high affinity binders, may also be applied to the selection of catalytic proteins. Therefore, we believe this method will have great potential for directed evolution of enzymes and the selection of new catalytic proteins.

Experimental Section

Synthesis of Biotinylated Ampicillin Sulfone. Ampicillin was obtained from Roche Diagnostics, and 4-methoxybenzyloxycarbonyloxyimino-2-phenyl-acetonitrile (MOZ-ON) and other chemicals were from Fluka. IR spectra were measured on a Perkin-Elmer 297 spectrometer; NMR spectra, on a Bruker ARX-300 (^{1}H) and a Bruker ARX-75 (^{13}C) spectrometer; and mass spectra, on a Perkin-Elmer Sciex API III$^+$ ESI-MS.

Ampicillin (1.5 g, 4.02 mmol) was dissolved in 4 mL of H_2O. One equiv of MOZ-ON (1.25 g, 4.02 mmol) in 4 mL dioxane was added.[49] The pH of the reaction mixture was adjusted to 7.8. After stirring for 6 h, the mixture was treated with 8 mL of H_2O (pH 7.8) and extracted with EtOAc (3×, 100 mL). The pH was adjusted to 3.0 with 10% H_3PO_4, and the aqueous phase was extracted with EtOAc (5×, 100 mL). The combined EtOAc extractions were dried with $NaSO_4$ and evaporated in vacuo. MOZ-ampicillin (1.58 g, 3.08 mmol, 76.5% yield) was collected as an egg white powder: IR ($CHCl_3$ [cm^{-1}]) 3405, 2960, 1785, 1725, 1695, 1612, 1515, 1495; ^{1}H NMR (300 MHz, DMSO): δ 1.42, 1.55, 3.74 (3× 3H, 3s), 4.23 (1H, s), 4.98−4.99 (2H, m), 5.38−5.41 (1H, m), 5.50−5.54 (2H, m), 6.91 (2H, d like m, J = 6.9), 7.24−7.35 (5H, m), 7.44−7.53 (2H, m), 7.88 (1H, br d, J = 7.9), 8.99 (1H, br d, J = 7.7); ^{13}C NMR (DMSO): δ 26.4, 30.2, 54.9 (3q), 57.4, 58.1 (2d), 63.5 (s), 65.4 (t), 67.1, 70.2 (2d), 113.6 (2C, d), 125.4 (d), 127.1, 128.0 (2× 2C, 2d), 128.6 (1C, s), 129.5 (d), 137.7, 155.6, 158.9, 168.9, 170.4, 173.2 (s); MS: 514.2 *m*/*z*.

(47) Jacob, F.; Joris, B.; Frere, J. M. *Biochem. J.* **1991**, *277*, 647−652.
(48) Mazzella, L. J.; Pazhanisamy, S.; Pratt, R. F. *Biochem. J.* **1991**, *274*, 855−859.
(49) Chen, S. T.; Wang, K. T. *Synthesis* **1989**, 36−37.

MOZ-ampicillin (513 mg, 1.00 mmol) was dissolved in 3.6 mL of cold H_2O, and the pH was adjusted to 7.0. $KMnO_4$ (190 mg, 1.2 mmol) was dissolved in 5 mL H_2O and treated with 65 μL 10% H_3PO_4. The $KMnO_4$ solution was added dropwise to the first solution on ice, while maintaining the pH between 6.8 and 7.3 with 10% H_3PO_4 and 10% NaOH. After 40 min, the reaction mixture was brought to pH 8.0 and filtered over Celite, and the filtrate was acidified to pH 3.0 with 10% H_3PO_4.[50] The product was extracted with EtOAc (5×, 100 mL, dried with $NaSO_4$, and concentrated in vacuo to afford 495 mg (91% yield) of MOZ-ampicillin sulfone as an egg white powder: IR ($CHCl_3$ [cm^{-1}]) 3410, 2960, 1810, 1730, 1700, 1612, 1515, 1495, 1328, 1118; ^{1}H NMR (DMSO): δ 1.33, 1.46, 3.74 (3 × 3H, 3s), 4.36 (1H, s), 4.97 (2H, s), 5.31 (1H, d, J = 4.5), 5.51 (1H, d, J = 8.3), 5.90 (1H, dd, J = 4.4, 8.8), 6.90 (2H, d, J = 8.6), 7.28−7.31 (5H, m), 7.35−7.50 (2H, m), 7.97 (1H, br d, J = 8.5), 8.65 (1H, br d, J = 8.8); ^{13}C NMR (DMSO): δ 17.4, 19.8, 55.1 (3q), 56.8, 59.1, 63.7 (3d), 64.6 (s), 65.4 (d), 67.3 (t), 113.8, 127.0 (2× 2C, 2d), 128.4 (s), 128.7 (d), 129.1, 129.9 (2× 2C, 2d), 136.0, 156.0, 159.5, 168.9, 170.6, 173.6 (6s); MS: 546.2 *m*/*z*.

A solution of MOZ-ampicillin sulfone (220 mg, 0.40 mmol) in 2 mL of CH_2Cl_2 and 0.2 mL of TFA was stirred at room temperature for 30 min.[49] The solvent was removed in vacuo. The product was triturated with ether that was removed as the supernatant after centrifugation. This procedure was repeated twice, and the product was dried in vacuo. Ampicillin sulfone (150 mg, 0.39 mmol, 97% yield) was isolated as a pale yellow powder: IR (KBr [cm^{-1}]): 3390, 2980, 1802, 1674, 1525, 1458, 1433, 1376, 1323, 1116; ^{1}H NMR (DMSO): δ 1.32, 1.42 (2× 3H, 2s); 4.33 (1 H, s); 5.31−5.29 (2H, m); 5.95 (1 H, dd, J = 8.4, 4.5,); 7.30−7.54 (5H, m); 9.12 (1H, d like m, J = 8.5); ^{13}C NMR (DMSO): δ 17.0, 19.1 (2q); 54.7, 55.6, 63.1 (3d); 63.7 (s), 64.6 (d), 127.4, 128.7 (2× 2C, 2d), 129.1 (d), 132.9 167.7, 168.0, 173.1 (4s), MS: 382.2 *m*/*z*.

EZ-LinkSulfo-NHS-LC-LC-Biotin (7.6 mg, 1.14 × 10^{-2} mmol, Pierce) was dissolved in 7 mL borate buffer pH 7.8. Ampicillin sulfone (7.1 mg, 2.3 × 10^{-2} mmol) was added. This solution was sonicated and stirred for 1 h at room temperature. Tris(hydroxymethyl)aminomethane base (2 mg, 1.7 × 10^{-2} mmol) was added, stirred for an additional 10 min, and lyophilized to afford 71.5 mg of a white powder (as a complex with borate) from which 30 mg was dissolved in MeOH and applied to a silica gel column (1 g, in a Pasteur pipet). The product was eluted with 10 mL of MeOH, which was removed in vacuo. The product was obtained quantitatively as a white powder from ampicillin sulfone. MS: 834.4 *m*/*z*.

Inhibition Studies. β-Lactamase (from *E. cloacae* (Fluka) or *E. coli* R-TEM β-lactamase from in vitro translation) in concentrations from 10^{-9} to 10^{-10} M (as determined by activity measurements,[51] and assuming that the specific activity of the in vitro translated β-lactamase is similar to that of the bacterially expressed enzyme) was incubated at room temperature in 50 mM potassium phosphate buffer (pH 6.0, 5% DMSO), with or without 1 mM ampicillin sulfone. DMSO is required to dissolve ampicillin sulfone. Aliquots of 500 μL were taken after 10, 20, 30, 45, and 90 min and were mixed with 500 μL of 0.4 mM nitrocefin (Becton Dickinson) in 50 mM potassium phosphate buffer (pH 6.0, 1% DMSO). Reaction kinetics were followed at OD_{486} during 2 min at room temperature with a Perkin-Elmer Lambda 20 spectrophotometer.[51] To show the covalent nature of the inhibition, a sample (500 μL) of inhibited and, as a control, not inhibited enzyme were dialyzed against 500 mL potassium phosphate buffer (50 mM, pH 7) in "Slide-A-Lycer 10K" devices (molecular weight cut-off, 10 kD, 0.5−3 mL sample volume) (Pierce) for 2 h before activity was measured.

Western Blot Analysis. *E. coli* S30-extract (as used for ribosome display [44]) was diluted 2-fold in 50 mM potassium phosphate buffer (pH 6.0, 1% DMSO) and incubated with 0.1 mM biotinylated ampicillin sulfone on ice. Aliquots were taken after various time points and excess inhibitor was removed by gel filtration (NAP-500 column, Pharmacia, according to the manufacturer's instructions). Alternatively, one aliquot was incubated at 25 C for 2 h while another was kept on ice at a pH 8 for 2 h. For inhibition studies, one aliquot was preincubated with 1 mM ampicillin sulfone for 1 h at room temperature prior to incubation with biotinylated ampicillin sulfone (as described above). The proteins were separated by SDS−PAGE (12%) and blotted onto a membrane (Immunoblot, Millipore). The membrane was blocked with milk, and the presence of the biotinylated suicide inhibitor was detected with an avidin−alkaline phosphatase conjugate (Pierce).

DNA Constructs for the Model System. The β-lactamase gene encoding the double cysteine to alanine mutant was PCR amplified from the plasmid pAP5__2C__2A [51] with the primers SDA-BLA 5 -(AGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATA TATCCATGGACTACAAAACCAGCCAGAAACGCTGGTGAAAGT), which codes for the Shine Dalgarno box and the beginning of the β-lactamase gene, and BLArev 5 -(ACACAGGCCCCCGAGGCCCAATGCTTAATCAGTGA), annealing at the end of the β-lactamase gene and introducing a unique *Sfi*I restriction site, while removing the stop codon. To create an inactive β-lactamase to use as a negative control for activity, the active site serine was mutated to alanine (S70A) by PCR mutagenesis. A unique *Msc*I restriction site was added at the site of the mutation to allow for discrimination of the genes encoding active and inactive enzyme. In parallel, a 171 amino acid portion of the *tol*A gene (residues 131 to 302) was amplified by PCR from chromosomal *E. coli* DNA in order to serve as a C-terminal tether for the displayed β-lactamase. The primers used were tolAforSFI 5 -(TATATGGCCTCGGGGGCCGAATTCCAGAAGCAAGCTGAAG), introducing an *Sfi*I-site, and tolAmrev 5 -(CCGCACACCAGTAAGGTGTGCGGTTAGCTCACCGAAAATATCATC), which introduced a RNA-stabilizing 3 -loop. The inactive S70A mutant and the active β-lactamase were fused to the *tol*A spacer by *Sfi*I digestion−ligation. The resulting constructs were further amplified by PCR with the primers tolAmrev and T7B 5 -(ATACGAAATTAATACGACTCACTATAGGGAGACCACAACGG), which introduced an RNA stabilizing 5 -loop and the T7 promoter for in vitro transcription. Both constructs were verified by DNA sequencing.

Separation of Ternary Complexes from Free β-Lactamase by Gel Filtration. For ribosome display, the in vitro transcription and in vitro translation of the constructs were carried out essentially as previously described.[44] Following a 10-min translation in 110 μL, the reaction was stopped by 4-fold dilution in ice-cold wash buffer (WB; 50 mM potassium phosphate buffer, pH 6.0; 150 mM NaCl; 50 mM $MgCl_2$). A 250-μL aliquot was treated with 2.5 μL RNase A (Qiagen, 10 mg/mL in WB) for 180 min at 4 C to release ribosome-bound β-lactamase, while another aliquot was treated with WB only. An aliquot (200 μL) of each solution was applied to a 1-mL gel filtration column (CL-4B, Pharmacia) that had been equilibrated with 10−20 mL of ice-cold WB. Fractions of 100 μL were collected, and the enzyme activity was assayed in a nitrocefin assay (200 μM nitrocefin in WB buffer, pH 6.0, 1% DMSO). In addition, an aliquot of diluted translation mix was treated with RNaseA for 1 h at room temperature without subsequent gel filtration. The β-lactamase activity of this sample was determined in order to quantify the total activity, as free enzyme, resulting from translation.

Enrichment for Catalytic Activity with Ribosome Display. RNA coding for active as well as inactive S70A mutant β-lactamase was mixed in a ratio of 1 to 10 (10 μg total of RNA in 10 μL water) and translated in 110 μL for 10 min.[44] Translation was stopped by 4-fold dilution into ice-cold WB with 1% BSA (bovine serum albumin, Sigma). Biotinylated ampicillin sulfone (stock solution, 0.01 M in DMSO) was added to a final concentration of 0.01 mM. A control sample contained the same final concentration of DMSO but no biotinylated ampicillin sulfone. After 30 min at 4 C, the reaction was stopped by removal of the excess biotinylated ampicillin sulfone by

(50) Johnson, D. A.; Panetta, C. A.; Cooper, D. E. *J. Antibiot.* **1963**, *28*, 1927−1928.
(51) Laminet, A. A.; Plückthun, A. *EMBO J.* **1989**, *8*, 1469−1477.

gel filtration (250 μL ribosomal complexes, 1 mL CL-4B column, as described above). Avidin-coated agarose beads (100 μL) (Roche Diagnostics), equilibrated in ice-cold WB containing 6% biotin-depleted sterilized milk,[44] were added to the flow-through (500 μL) and gently shaken at 4 C for 30 min to capture the labeled complexes. The beads were then washed five times with 500 μL of ice-cold WB. During this procedure, the beads were transferred to a new prechilled tube. To elute the mRNA, 250 μL EB20 buffer (50 mM Tris-HCl, pH 7.5, 20 mM EDTA) was added, and the solution was gently shaken for 10 min at 4 C before centrifugation (2 min, 2000*g*). The supernatant (200 μL) was used for mRNA purification (High Pure RNA Isolation Kit, Roche Diagnostics). Reverse transcription (RT) and PCR was carried out as described,[44] except that the primers used were SD-BLA and BLArev.

Monitoring the Enrichment of Active β-Lactamase Relative to the Inactive S70A Mutant. The product from RT-PCR was purified with a spin column (QIAquick, Qiagen) according to the manufacturer's protocol. The DNA was subsequently analyzed by restriction analysis with *Msc*I or by sequencing. After digestion with *Msc*I, the ratio of active (uncleaved) to mutant (cleaved) β-lactamase was estimated from the relative intensity of bands on a 2.5% Metaphor agarose gel (FMC) using the program NIH image. DNA sequencing was carried out with 10 ng PCR product according to the manufacturer's protocol (SequiTherm EXCEL *II*, Epicenter Technologies). The enrichment factor was estimated by determining the relative signal intensities of the 2 nucleotides that were mutated in the inactive S70A enzyme, with NIH image.

β-Lactamase Inhibitory Protein (BLIP) Expression as pD Fusion. We constructed a fusion protein of BLIP with protein D, carrying an N-terminal His_6 tag (gpHD). The BLIP (Swiss-Prot P35804) gene was amplified by PCR and cloned into the vector pAT39 [43] between the *Bam*HI and *Hind*III sites to yield gpHDBLIP, consisting of gpHD, a Gly-Ser linker (GSGGGSGGGG), and BLIP. The fusion protein was amplified from pAT39 by PCR and ligated into the vector pET40-b (Novagen), carrying a kanamycin resistance gene. Plasmids were verified by sequencing and subsequently used for expression in the *E. coli* strain BL21 (DE3) [pLysS] (Stratagene) in 1 L of LB medium containing 100 μg/mL kanamycin. Expression was induced with IPTG (1 mM) at $OD_{600} = 0.8$. The cells were grown for 3 h at 37 C and harvested by centrifugation. The cell pellet was resuspended in PBS (10 mM sodium phosphate buffer, pH 7.4, 140 mM NaCl, 15 mM KCl), and the cells were lysed by sonication. The resulting crude extract was centrifuged (5 min, 6000 g, 4 C), and the pelleted inclusion bodies were washed once with PBS and resuspended with a solution containing 6 M GdnHCl, 0.1 M Tris, pH 8.0, and 10 mM imidazole. The solubilized inclusion bodies were applied to a Ni−NTA column (15 mL) (Qiagen) and washed with GdnHCl buffer. The protein was refolded on the column by replacing the GdnHCl-buffer with refolding buffer (20 mM Tris pH 8, 500 mM NaCl, 20% glycerol, GSH (0.2 mM)/GSSG (1 mM)) in a linear gradient and eluted with 250 mM imidazole in refolding buffer.[52] The protein was dialyzed against PBS and concentrated to 2.5 mg/mL. One liter of *E. coli* culture yielded 60 mg of gpHDBLIP, which was frozen in liquid nitrogen and stored at −80 C. The protein gave a single band in SDS−PAGE upon staining with Coomassie-Blue. The β-lactamase inhibitory activity of gpHDBLIP was quantified in β-lactamase inhibition experiments as described for ampicillin sulfone (data not shown).

Comparison of Affinity Selection and Selection for Catalytic Activity. Ribosome display was carried out as described above, but only with active β-lactamase. After stopping the translation, the stopped mix was incubated for 4 h on ice. A 200-μL aliquot was treated for 1 h on ice with biotinylated ampicillin sulfone (final concentration, 0.01 mM), while another aliquot was treated with gpHDBLIP (final concentration, 0.3 μM). The selection procedure for the aliquot treated with biotinylated ampicillin sulfone was as described above. A gel filtration step was performed, as described for the fraction with the suicide inhibitor. The ribosomal complexes binding to gpHDBLIP were captured via the His_6 tag by incubating the flow-through with 10 μL of Tetra-His antibody (Qiagen, 0.1 μg/μL, in WB) and 100 μL of protein G PLUS-Agarose (Santa Cruz) equilibrated in WB in the presence of 4% sterilized milk powder. The agarose was washed as described for the magnetic beads. Rescuing of the mRNA, RT, and PCR was carried as out described above.

Abbreviations: RT, reverse transcription; PCR, polymerase chain reaction; GSH and GSSG, the reduced and oxidized forms of glutathione, respectively; GdnHCl, guanidine hydrochloride; bp, base pairs; BLIP, β-lactamase inhibitory protein; *E. coli*, *Escherichia coli*; GFP, green fluorescent protein; gpHD, His-tagged gene product D from phage lambda; IPTG, isopropyl-β-D-thiogalactoside; PAGE polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; MOZ-ON, 4-methoxybenzyloxycarbonyloxyimino-2-phenyl-acetonitrile.

Note Added in Proof: After this paper was submitted, it was reported (Takahashi, F.; Ebihara, T.; Mie, M.; Yanagida, Y.; Endo, Y.; Kobatake, E.; Aizawa, M. *FEBS Lett.* **2002**, *514*, 106−110) that 4 random point mutants of dihydrofolate reductase, which had been obtained in a ribosome display selection for binding to the inhibitor methotrexate, were active. This suggests that enrichment of active enzyme variants may also be possible simply by binding to an inhibitor under some circumstances, not requiring the turnover-based selection reported here, even though inactive mutants that still bind to an inhibitor are certainly conceivable.

Acknowledgment. We thank Prof. Dr. Manfred Hesse for the possibility to perform the organic synthesis in his laboratory facilities. We also thank Patrik Forrer, Michael Stumpp, Kaspar Binz, and Stephen Marino for fruitful discussions and technical advice. This work was supported by Schweizerischer Nationalfonds Grant 3100-046624.96/1. Joelle Pelletier was a recipient of a fellowship from le Conseil de Recherche en Sciences Naturelles et en Génie du Canada.

JA025870Q

(52) Holzinger, A.; Phillips, K. S.; Weaver, T. E. *BioTechniques* **1996**, 20, 804−806, 808.

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 279, No. 18, Issue of April 30, pp. 18870–18877, 2004

Printed in U.S.A.

Directed *in Vitro* Evolution and Crystallographic Analysis of a Peptide-binding Single Chain Antibody Fragment (scFv) with Low Picomolar Affinity*

Received for publication, August 19, 2003, and in revised form, January 6, 2004
Published, JBC Papers in Press, January 30, 2004, DOI 10.1074/jbc.M309169200

Christian Zahnd‡, Silvia Spinelli§¶, Béatrice Luginbühl‡¶, Patrick Amstutz‡, Christian Cambillau§, and Andreas Plückthun‡‖

From the ‡Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland and §Architecture et Fonction des Macromolécules Biologiques, CNRS, 31 Chemin Joseph Aiguier, F-13402 Marseille Cedex 20, France

We generated a single chain Fv fragment of an antibody (scFv) with a binding affinity of about 5 pM to a short peptide by applying rigorous directed evolution. Starting from a high affinity peptide binder, originally obtained by ribosome display from a murine library, we generated libraries of mutants with error-prone PCR and DNA shuffling and applied off-rate selection by using ribosome display. Crystallographic analysis of the scFv in its antigen-bound and free state showed that only few mutations, which do not make direct contact to the antigen, lead to a 500-fold affinity improvement over its potential germ line precursor. These results suggest that the affinity optimization of very high affinity binders is achieved by modulating existing interactions via subtle changes in the framework rather than by introducing new contacts. Off-rate selection in combination with ribosome display can evolve binders to the low picomolar affinity range even for peptide targets.

Directed evolution of proteins *in vitro* has become a widely applied strategy to generate proteins with a desired property (1). Especially in the generation of high affinity binders, the iterative succession of randomization and selection was shown to efficiently mimic natural affinity maturation. The success of this approach is dependent on the size and the quality of the library and the power of the selection method used. Technologies that work entirely *in vitro* such as ribosome display (2) and mRNA display (3) are more favorable than methods that work partially *in vivo* such as phage display (4) or fully *in vivo* such as the yeast two-hybrid system (5) or the protein complementation assay (6) as the *in vitro* technologies do not need transformation of cells after each new round of randomization. Therefore, they allow the handling of much larger libraries and more cycles of randomization, and the experimental work is accelerated greatly.

A prerequisite of any evolution is randomization between different selection rounds. In ribosome display this is facilitated by the use of linear DNA. The randomization occurs at a low rate by the intrinsic error rate of the polymerase used but can be enhanced by error-prone PCR (7), by DNA shuffling (8), or both, thereby generating highly diverse pools.

While the generation of binders having binding constants in the subnanomolar range can be achieved, *e.g.* with synthetic antibody libraries and established techniques (9), the generation of very high affinity binders with binding constants in the lowest picomolar affinity range is difficult for several reasons. First, a very stringent selection pressure must be applied to separate improved binders from the already very high affinity precursors. Second, selected binders must be eluted efficiently, which may become very difficult for binders with very slow off-rates. Here ribosome display offers a significant advantage since bound binders must not be eluted, but the ribosomally bound mRNA can be recovered by the addition of chelating agents, which destabilize the ribosomal complex (1). In particular the generation of very high affinity peptide binders is made difficult by the relatively high flexibility of the peptide in the unbound state and the corresponding loss of entropy upon binding. This is less of a problem for comparatively rigid antigens such as hydrophobic small molecular weight compounds for which subpicomolar binders are known (10).

We applied a competitive selection for increased off-rates to affinity mature a high affinity peptide binder previously selected with ribosome display from a murine library (11). The peptide was derived from the yeast transcription factor GCN4. We constructed different mutants of a high affinity binder and generated from them second generation libraries using DNA shuffling and error-prone PCR. From these libraries we successfully isolated binders in the low picomolar affinity range. By determining the crystal structure both in the free and antigen-bound state we could show that the gain in affinity of 500-fold, compared with its likely germ line precursor, was almost exclusively a result of second sphere mutations not being in direct contact with the antigen. These findings may have great impact on future library design and affinity maturation strategies.

EXPERIMENTAL PROCEDURES

Expression and Purification of Single Chain Fv Fragments—The scFv[1] genes were cloned into the periplasmic expression vector pAK400 (12) introducing a His_6 tag, expressed in *Escherichia coli* SB536 (13),

* This work was supported by Schweizerische Nationalfonds Grant 31-65344.01. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1P4I and 1P4B) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ Both authors contributed equally to this study.

‖ To whom correspondence should be addressed. Tel.: 41-1-6355570; Fax: 41-1-6355712; E-mail: plueckthun@bioc.unizh.ch.

[1] The abbreviations used are: scFv, single chain Fv fragment; V_L, variable domain of the light chain; V_H, variable domain of the heavy chain; 8-oxo-dGTP, 8-oxo-2′-deoxyguanosine 5′-triphosphate; dPTP, 6-(deoxy-β-D-*erythro*-pentofuranosyl)-3,4-dihydro-8*H*-pyrimido-[4,-5*c*][1,2]oxazine-7-one-5′-triphosphate; RIA, radioimmunoassay; CDR, complementarity-determining region.

TABLE I
Crystal parameters and refinement statistics

	Free scFv	Complex
Data collection		
Space group	$MP2_12_12_1$	$P2_1$
Total/unique number of reflections	51,966/6,870	129,374/10,027
Percentage of data $>1\sigma$ (overall/last shell)[a]	96.0/96.0	97.2/97.6
Overall I/σ (I) (last shell)	11.9/2.6	8.5/4.1
Resolution limits	24.6/2.8	20.0/2.3
R_{merge} (%) (overall/last shell)[a]	5.9/28.1	6.2/18.4
Refinement		
Number of protein/solvent atoms	1,658	17,648/66
Number of reflections	6,796	9,940
Resolution limits (Å)	12.0–2.8	15.0–2.35
R/R_{free} value (%)	20.05/26.4	18.0/21.5
r.m.s.d.[b] on bonds (Å)/angles (°)	0.012/2.0	0.025/2.1
r.m.s.d. on impropers/dihedrals (°)	1.14/27.8	0.86/25.2
Mean B-factor (Å^2) (main/side chain)	35.4	23.8/27.0
Peptide (main/side chain)		25.0/28.0
Water		33.0

[a] Last shell, 2.9–2.8 Å (free scFv); 2.43–2.35 Å (complex).
[b] r.m.s.d., root mean square deviation.

and purified by immobilized metal ion affinity chromatography and subsequent antigen affinity chromatography as described previously (11). For structure determination, a seleno-Met-containing variant of clone H6 was grown in 1 liter of M9 minimal medium to an OD_{600} of 0.6. An amino acid mixture containing 100 mg/liter each of Lys, Thr, and Phe and 50 mg/liter each of Leu, Ile, Val, and seleno-Met was added. After 1 h, isopropyl-1-thio-β-D-galactopyranoside was added to a final concentration of 1 mM for expression overnight. The protein was purified as described above.

The Library Construction—The scFv fragments C11L34, L24, L107, L135, L107–135, H6, and H67 in the vector pAK400 were PCR-amplified using primers SDAla+ (5′-AGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATCCATGGCGGACTACAAAGAT) and Sfi_rescue (5′-GCCCTCGGCCCCCGAGGC). A total of 1 μg of PCR product of an equimolar mixture of all clones was used for DNase I shuffling (14) as described previously (15). Some of the reassembled PCR products were further randomized by error-prone PCR using primers SDAla+ and Sfi_rescue. Error-prone PCR was performed using the dNTP analogues 8-oxo-dGTP and dPTP according to Ref. 7 with small modifications. Twenty-six cycles of error-prone PCR were performed in the presence of 85 μM dPTP, 85 μM 8-oxo-dGTP, and 50 μM dGTP, dATP, dTTP, and dCTP each. The final mutation rate after DNase I shuffling and error-prone PCR and the distribution of the mutations were determined by sequencing about 2000 bp. A gene III linker was fused to the library as described earlier (11).

Selection for Improved Affinities—The library was transcribed and translated *in vitro* as described in Ref. 11. The ternary complexes of ribosome, mRNA, and displayed proteins were equilibrated with 1 nM biotinylated peptide GCN4(7P14P) (16) at 4 °C overnight. Every sample was split into two aliquots, and only to one, 1 μM non-biotinylated GCN4(7P14P) was added. The aliquots were incubated for a defined time span, which was increased from round to round, starting with 2 h in the first round and going up to 10 days after the fourth round in a rollover shaker at 4 °C. The complexes were recovered by binding to streptavidin-coated magnetic beads (Roche Applied Science) for 30 min. The beads were washed five times, and the RNA was eluted and purified as described in Ref. 11.

Affinity Comparison of Pools and Single Clones by Radioimmunoassay (RIA) and Inhibition BIACORE—For the analysis of single clones, the selected pools were cloned into pTFT74 (17). RNA of single clones was transcribed directly from plasmid pTFT74, whereas the pools were transcribed from a PCR product. RIAs were performed as described previously (18). Of some clones, absolute affinities were measured using the inhibition method on a BIACORE 3000 (19). The purified protein was diluted to 1 nM and incubated with different concentrations of antigen overnight at 4 °C. The samples were injected over a chip, which was coated to maximal density with the antigen used for selection. The slope of the association curves in the linear phase was plotted against the concentration of soluble antigen. K_D was determined from at least three independent curves as described previously (11).

Crystallization—Crystals of clone H6 in complex with the antigen and clone C11L34 in the absence of the peptide were obtained using the hanging drop method. Drops of 1 μl (protein concentration, 8 mg/ml) of the unliganded Fv GCN4 were mixed with 1 μl of the well solution (1.1 M ammonium sulfate, 150 mM sodium citrate, pH 4.8). Crystals appeared after 2 days at 20 °C belonging to space group $P2_12_12_1$ (a = 35.08 Å, b = 60.53 Å, and c = 123.05 Å) and contained one molecule per asymmetric unit (V_m = 2.29 Å^3/Da, 46% solvent (20)). Crystals of the complex of H6 and the peptide YHLENEVARLKK were obtained by mixing the scFv and the peptide in a 1:2 molar ratio. Drops of 1 μl of the complex (protein concentration, 7.4 mg/ml) were mixed with 1 μl of a solution containing 32–28% (w/v) polyethylene glycol 4000, 0.1 M Tris/HCl, pH 7.5. The space group was $P2_1$ (a = 37.24 Å, b = 36.29 Å, c = 84.46 Å, β = 90.5°) with one complex per asymmetric unit and a specific protein volume of 2.03 Å^3/Da.

Data Collection and Processing—Data of the free scFv fragment were collected at 100 K on a MAR-Research Imaging Plate (MAR, Hamburg, Germany) placed on a Rigaku RU200 rotating anode using the CuK_α radiation. Crystals were frozen using 25% glycerol, and they diffracted to 2.6 Å. Data integration and reduction were performed using DENZO (21) and SCALA (22). A crystal of the complex was frozen with 16% glycerol. Data were collected at 100 K in the beam-line ID14-EH4 at the European Synchrotron Radiation Facility (Grenoble, France), λ = 0.988 Å. The data reduction was performed with SCALA (22). The data of the complex were merged with TRUNCATE (22) in CCP4i using the anisotropic correction option. The R_{Sym} for the free scFv and the complex were 5.9 and 6.2%, respectively. The statistics of the data sets are shown in Table I.

Structure Determination and Refinement—Both structures were solved by molecular replacement (23) using the program AMoRe (24). Rotation and translation searches were performed using the murine Fv SE155-4 (Protein Data Bank entry 1MFA) for the unliganded Fv. The structure of the unliganded Fv was used to solve the structure of the complex. The molecular replacement solution for the unliganded Fv gave an initial correlation coefficient of 45% and R-factor of 47%, while the values for the complex were 46 and 48%, respectively, between 8.0 and 3.0 Å. Refinement was performed first with CNS (25) using standard protocols followed by REFMAC (26) using maximum likelihood, incorporating bulk solvent corrections and translation-libration-screw (TLS) anisotropic refinement. After each refinement cycle, a new map was calculated, and the model was fitted with Turbo-Frodo (27). Final refinement data are summarized in Table I. The coordinates and structure factors of the free and peptide-bound Fv have been deposited in the Protein Data Bank at Research Collaboratory for Structural Bioinformatics as entries 1P4I and 1P4B, respectively.

RESULTS

Library Construction—In a previous study, we selected a group of scFv fragments from a murine library by using ribosome display, a selection method that works entirely *in vitro* and therefore allows the selection from very large libraries (Fig. 1). The selected scFv fragments bound with very high affinity to the peptide GCN4(7P14P) derived from the yeast transcription factor GCN4 (11, 16). The highest affinity clone, named "C11L34," had an affinity of 40 pM and acquired a crucial amino acid substitution during ribosome display rounds

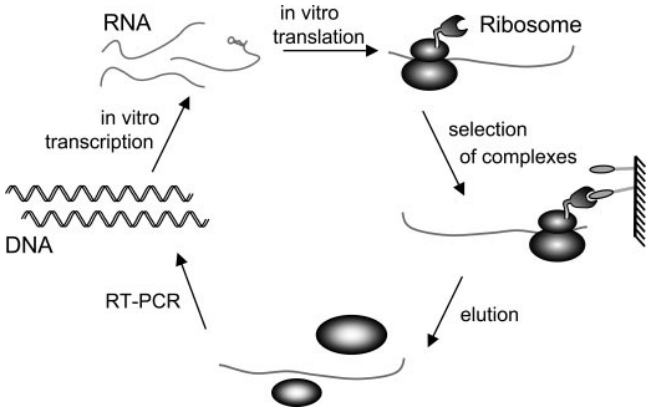


FIG. 1. **Ribosome display.** Ribosome display is a method to select binders from large libraries. A library of mRNA molecules, encoding for potential binders, is translated with an approximately stoichiometric amount of ribosomes *in vitro*. Since the mRNA contains no stop codon and the reaction is stopped with a buffer containing high concentrations of magnesium, the ribosome will not dissociate and form a stable complex with the translated protein and the mRNA. A spacer fused to the C terminus of the library takes up the place in the ribosomal tunnel and thereby ensures that the freshly translated protein can fold into its native structure even if it is displayed on the ribosome. The complexes are used for the selection of binders to immobilized antigen. After washing, the mRNA is eluted using chelating agents such as EDTA to dissociate the ribosome. Eluted mRNA is reverse transcribed, amplified by PCR, and potentially randomized to introduce more diversity. mRNA for the next selection round is transcribed directly from a PCR product. Note that at no step does the library have to be transformed into cells, which would otherwise drastically reduce the size of the library that is available for screening. The PCR can also be carried out under error-prone conditions thereby allowing the library to evolve (for details, see text). *RT*, reverse transcription.

that led to a 65-fold affinity improvement compared with its likely progenitor from the murine B-cells (11).

We constructed a second generation library in two steps. First, we generated five different point mutants of the high affinity clone and a double mutant, based on mutations that had been enriched in the first selection experiment, and measured their affinities (Table II and see Fig. 4). These mutants were taken as a template for the construction of three second generation libraries. The first pool, named "S," was generated by using DNA shuffling of all seven clones including C11L34, thereby recombining all mutations (8). Two other pools were generated by the use of error-prone PCR. Pool "R" was generated by error-prone amplification of C11L34, and pool "SR" was generated by the combination of the two approaches: the mutants were amplified by error-prone PCR and then subjected to DNA shuffling.

To enhance the mutation rate of the polymerases, we used high concentrations of the nucleotide analogues 8-oxo-dGTP and dPTP leading to mismatch incorporation (7). The final mutation rates of the pools were determined as 8.9 kbp^{-1} for pool S, 61 kbp^{-1} for pool R, and 78 kbp^{-1} for pool SR. Taking into account that many mutations are silent or have no effect on binding, *e.g.* because they are located in the linker or the lower framework, we estimated that the randomized pools will have two to three relevant amino acid substitutions per gene. As a consequence of the PCR-based randomization we found that mutations that were introduced during an early PCR cycle were more prominent in the final library. This clustering of mutations could be circumvented in future experiments by the use of a high template concentration and very high error rates, thereby reaching the final mutation rate within fewer amplification cycles.

Ribosome Display and Off-rate Selection—For ribosome display, all three libraries had to be fused to a protein spacer derived from gene III to allow the displayed protein to fold properly (15). Since panning followed by extensive washing would hardly be able to discriminate between binders with different affinities all lying in the picomolar range, we used a competitive selection for decreased off-rates. The ribosomal complexes formed after *in vitro* translation were equilibrated overnight with a 1 nM solution of biotinylated antigen. A 1000-fold excess of competitor antigen carrying no biotin label was then added, and the pools were incubated at 4 °C. All complexes dissociating from the biotinylated antigen, to which they were initially bound, will be captured by the competitor carrying no biotin label and thus cannot be bound to streptavidin-coated magnetic beads. Hence the duration of incubation with competitor is defining the stringency of the selection. The incubation was prolonged from round to round (2, 10, and 240 h), thereby increasing the selection pressure. After every round, selected mRNA was isolated and reverse transcribed, and the enriched pools were subjected to DNA shuffling. The pool SR was further randomized after the second round using the same conditions as in round 1.

Interestingly the mRNA of pool R, which had been generated by error-prone amplification of C11L34 only but which had not been subjected to DNA shuffling, could not be restored after the first round of selection. This indicates the importance of recombination in conjunction with high mutation rates to preserve a fraction of the pool in an active form. Furthermore the long off-rate selection times underscore the stability of the non-covalent ribosomal complex, which can survive more than 20 days at 4 °C.

Analysis of the Pools after Off-rate Selection—After every round of ribosome display, the pools were checked for improved binding by RIA (18). The pools were expressed *in vitro* in the presence of [^{35}S]Met, equilibrated with different amounts of free antigen, and allowed to bind to surface-immobilized antigen. The amount of competitor antigen needed to inhibit the binding of the scFv fragments to surface-immobilized antigen correlates with the mean affinity of binders found in the pool under investigation and decreased from round to round (Fig. 2). In the initial error-prone randomized pool even high concentrations of competitor did not affect binding of the pool to the plate. After 240 h of off-rate selection, however, 0.1 nM antigen was sufficient to prevent 50% of the pool from binding to the surface compared with the uninhibited signal, giving evidence that the mean affinity of the binders in the pool had improved affinities compared with clone C11L34. The total signal intensity of the pools decreased from round to round, indicating that the percentage of rescued binders decreased from round to round due to the very stringent selection pressure.

Screening for Binders—After the third round, pools SR and S were cloned, plasmids of single colonies were isolated, and *in vitro* expressed protein was analyzed by RIA. Only 7 of 54 clones (13%) showed a significant binding signal to the surface-immobilized antigen in the absence of competitor. All of these clones could be completely inhibited with 10 nM antigen. Four of them were inhibited at even lower concentrations than C11L34 and were therefore ranked as affinity improved. In pool S, which was generated by DNA shuffling, 4 of 15 (25%) clones analyzed showed binding to the antigen after the third round of directed evolution, but only one had an improved inhibition signal (Table II).

Only a few molecules had reached improved affinities that let them survive the applied selection pressure, whereas the background signal, consisting of unspecific complexes and RNA sticking to the streptavidin-coated magnetic beads, remained constant. To improve the ratio of binders over background, a non-stringent enrichment round was performed. After this

TABLE II
Summary of the mutations and affinities of different GCN4 binders

Clone[a]	Mutations[b]		K_D[c]
			pM
C11L34[d]	L42(Asn → Ser)		40 ± 4
L24	L24(Arg → His)		32 ± 13
L107	L107(Ala → Val)		48 ± 6
L135	L135(Asn → Asp)		23 ± 4
H6	H6(Glu → Gln)		20 ± 2
H67	H67(Ile → Val)		47 ± 7
L107L135	L107(Ala → Val)	L135(Asn → Asp)	47 ± 5
L135H6	L135(Asn → Asp)	H6(Glu → Gln)	16 ± 13
52SR4	L13(Thr → Ser)	L107(Ala → Val)	5.2 ± 2.3
	L135(Asn → Asp)	H30(Ser → Leu)	
63S3	L145(Leu → Pro)		ND[e]
70SR4	L13(Thr → Ala)	L135(Asn → Asp)	ND
	H96(Leu → Pro)		
82SR4	H70(Tyr → His)	H94(Asn → Ser)	ND
84SR4	H32(Thr → Ala)	H70(Tyr → His)	ND
	H94(Asn → Ser)		

[a] The first eight clones listed were found in a previous selection (11), and their mutations compared with C11L34 were introduced by site-directed mutagenesis. Clones H6 and L135H6 were constructed because it was shown recently that the residue on position H6 is often critical for stability and affinity (32, 34). All these clones were used for the generation of different libraries by DNA shuffling and error-prone PCR with the exception of clone L112H6, which was constructed at a later point in the project. The last five clones listed were found during the directed evolution and showed promising signals in a competition RIA.

[b] The mutations are given in the AHo nomenclature (35).

[c] The affinities were determined by inhibition BIACORE. The affinity improvement during the directed evolution was monitored mainly by competition RIA (see text).

[d] Clone C11L34 was found to have a single point mutation compared with its likely progenitor (11). All other clones are derived from C11L34 and contain its mutation, L42(Asn → Ser). For historical reasons, its name is indicated in the Kabat nomenclature (36) and was retained.

[e] ND, not determined.

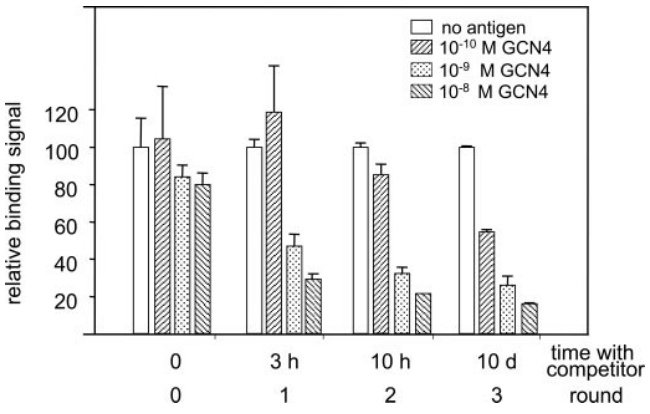


FIG. 2. **Inhibition RIA of pools after subsequent rounds of off-rate selection.** Three rounds of ribosome display were performed using selection for improved off-rates by competing biotinylated peptide with unlabeled peptide (see text). After each round, ^{35}S-labeled protein of the pools was expressed *in vitro* and equilibrated with different amounts of antigen. The equilibrated pools were then bound to surface-immobilized antigen. The binding signal of the pool is a measure of the amount of free scFv fragment present at equilibrium. High affinity binders are completely complexed by small concentrations of free antigen and thus do not bind to the immobilized antigen. The less free antigen is needed to prevent a pool from binding to the surface-immobilized antigen, the higher the mean affinity of the pool is. It can be seen that the concentration of free antigen needed to prevent the pool from binding to surface-immobilized antigen decreases from round to round. *d*, days.

enrichment, 14 of 16 (87%) randomly picked clones showed binding to the antigen, of which eight showed improved inhibition patterns compared with C11L34. Furthermore the signal intensity of the pool in the RIA increased by a factor of 120, indicating a strong enrichment of the binders in the pool over the nonspecific RNA. Thus, a non-selective enrichment round after extensive off-rate selection may be useful in general to amplify the selected clones.

Sequences of the Clones with Improved RIA Signal—The clones showing the most promising RIA signal were sequenced. They carried an average mutational load of one to four amino acid substitutions, whereas zero to two mutations derived from shuffled input DNA (Table II). The mutations were distributed over both domains, and some mutations showed up several times. It is likely that they were found independently since they had different codon usage. Interestingly the only mutation lying in CDRs, L135(Asn → Asp), originated from the clones used for DNA shuffling. All other mutations were located in framework positions.

BIACORE Measurements—The affinity of all clones generated by site-directed mutagenesis and used for the library construction was measured. In addition, the binding constant of the evolved clone showing the best RIA signal after off-rate selection was determined. All clones were expressed in the periplasm of *E. coli* and purified by immobilized metal ion affinity chromatography and antigen affinity chromatography (18). The dissociation constant K_D of the purified proteins was determined in solution by competition BIACORE analysis (19, 28, 29).

The dissociation constants of the clones used for the library generation were between 20 and 50 pM (Table II). Clones L135(Asn → Asp) and H6(Glu → Gln) showed a significantly improved affinity of 23.3 and 20 pM, respectively. Therefore, we also constructed the double mutant L135H6. It had an affinity of 16 pM. However, this clone had not been available when the library was constructed. The affinities of the evolved clones were mainly monitored by RIA. The affinity of the clone with the best RIA signal, named 52SR4, was determined. It was found in pool SR after the third round, and its affinity in solution was determined to be 5.2 pM by the competitive BIACORE method (11). All measurements were repeated three times independently, resulting in uncertainties of 5–45%. Measuring affinities in this range is very difficult. The minimal concentration of the scFv that could be detected using BIACORE was about 1 nM. Due to the very high affinities, this is much above the K_D value at which concentration the measurement should have been performed ideally. However, even if the K_D of this highest affinity binder cannot be given with satisfy-

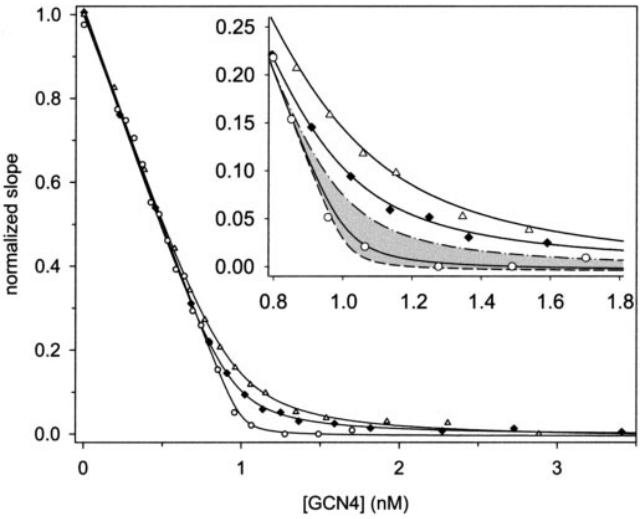


FIG. 3. **Affinity determination using inhibition BIACORE.** The dissociation constant K_D in solution of different single chain Fv fragments using inhibition BIACORE was determined at 20 °C. The *triangles* correspond to the starting clone C11L34, having a K_D of 40 pM, the *diamonds* correspond to clone H6 from which the crystal structure in complex with the antigen was determined to have a K_D of 20 pM, and the *circles* correspond to 52SR4, which was the evolved clone with the best RIA signal having a K_D of 5.2 pM. For better comparison, the theoretical curves for a 10 pM affinity (*dash-dotted line*) and a 1 pM affinity (*dashed line*) are indicated in the *inset*. Due to the very high affinities the measurements become close to the detection limit. The given value of 5.2 ± 2.3 pM was obtained in three independent measurements.

ing accuracy (5.2 ± 2.3 pM, determined from several independent protein preparations), a clear improvement of the scFv over the initial constructs can be observed (Fig. 3). The K_D values of C11L34 and 53SR4 were also determined by kinetic BIACORE analysis with very low coating density, and the equilibrium K_D data were confirmed (data not shown).

The Overall Structure of the scFv—The crystal structure of C11L34 was determined in the absence of the antigen to 2.35 Å, and the crystal structure of variant H6 in the presence of the antigen was solved to 2.8-Å resolution (Fig. 4*A*). The C-terminal His tag and the linker connecting the two variable domains were not defined in the electron density as is typical for other scFv fragments. The N-terminal part of V_H of the scFv in complex with the peptide was poorly defined. Initial attempts to crystallize the complex with the 33-amino acid peptide used for selection were not successful. Previous studies with CD had suggested that the peptide would adopt a random coil conformation (16) in the antibody-bound form, which implied that not more than 10–12 amino acids of the extended peptide would be recognized by the scFv. Therefore, we used a truncated peptide of 12 amino acids for a second crystallization trial. With this approach crystals were obtained. When the structure was solved, the peptide was found to bind in a three-turn helical conformation, thereby making its total contact area smaller than expected. Therefore, not all of the potentially interacting residues can be seen in the present structure. To better illustrate the complete potential binding interface, the peptide was elongated at its N terminus, and the corresponding residues, which we denote, M1 to M5, in contrast to the structurally resolved residues of the peptide, P1 to P12, were modeled in an α-helical conformation into the complex. We repeated the CD measurements, which led to the original assumption that the bound peptide would adopt a random coil conformation (16). From the spectra of the scFv alone compared with the scFv bound to the peptide we attempted to deduce the spectrum of the peptide when bound to the scFv. Whereas the peptide in solution was clearly random coil, the spectrum of the peptide in complex with the scFv, originally interpreted as random coil, was very noisy and did not allow a clear conclusion (data not shown). The antibody had originally been raised against a variant of the coiled-coil leucine zipper of the yeast transcription factor GCN4 in which two residues were changed to proline (16). The origin of the peptide explains that there is some residual tendency for α-helix formation. Indeed the almost perfect helical conformation of the peptide in complex with the scFv suggests that the helical conformation is further induced by binding to the antibody.

Conformational Changes Due to Antigen Binding—Superposition of the Fv structure in the antigen-bound and free state results in a root mean square deviation of only 0.75 Å. Deviations are located between residues H9–18 and H48–52, and they are mostly due to the poor definition of the N-terminal part of V_H in the structure of the complex. These changes located opposite to the binding site are probably not related to the complex formation. A major rearrangement, however, was seen in CDR H3. Specifically residues Gly-H109 and Leu-H110 deviate strongly by 3.4 and 1.4 Å, respectively. Since this segment is well defined in both electron density maps, it is clear that the conformational change is directly induced by the binding of the peptide to the scFv, namely by Arg-P9 (Fig. 4*B*). Beside these changes, only small structural changes in the CDRs are observed upon peptide binding, which are all less than 1 Å. Upon peptide binding, the two domains of the scFv undergo a small rotation relative to each other by 1.25°. This may allow a more optimized binding geometry (see below).

The Interaction of the scFv with the Antigen—The scFv forms a deep (6–8 Å) and broad (8 Å) cleft of 20 Å in length to which all six CDR sequences contribute. The bottom of the cleft is formed mainly by the relatively short CDR3 loops of the heavy and light chain. The borders are formed by the loops of CDR1 and CDR2 of both chains. The antigen lies as an almost ideal three-turn α-helix in the binding cleft. This engulfed binding of the antigen results in a buried surface of 700 Å^2, which corresponds to 45% of the total water-accessible surface of the peptide. Modeling the peptide in its full length into the binding groove by extending the helix yields in an interaction interface of about 1580 Å^2, which is bigger than the interface between the light chain and the heavy chain (1479 Å^2). This buried surface value is also larger than those observed between most Fab fragments/Fv fragments in complex with proteins (30).

The interaction is made up by a great number of specific contacts between the scFv and exclusively the side chains of the peptide. The helical peptide faces the scFv with the same side that is responsible for the leucine zipper formation in the natural protein (31). Almost 30% of the surface of the peptide is contributed by His-P2. Its aromatic plane forms a hydrophobic stacking interaction with the indole moiety of Trp-H59. His-P2 also establishes a hydrogen bond to Asp-H65 and is further stabilized by a hydrophobic contact to Ile-H67. The neighboring residue Leu-P3 is held between three aromatic residues (Tyr-L40, Trp-L109, and Trp-L137), and Arg-P9 on the adjacent helical turn makes a strong (double) ionic interaction with Asp-H137. The interaction with Arg-P9 enforces the conformation of CDR H3 (H109–137), especially that of Gly-H109. The conformation of Arg-P9 is most likely locked by a hydrogen bond to Glu-P6 (Fig. 4*B*). In addition, Glu-P6 establishes a contact to the backbone carboxyl oxygen of Gly-H40 via a water molecule. An ionic hydrogen bond is also made between Lys-P4 and Asp-H69. A great number of hydrophobic interactions, namely stacking of aromatic rings and packing of aliphatic side chains, are found. The pattern of charged and hydrophobic residues observed on the surface of the peptide is comple-

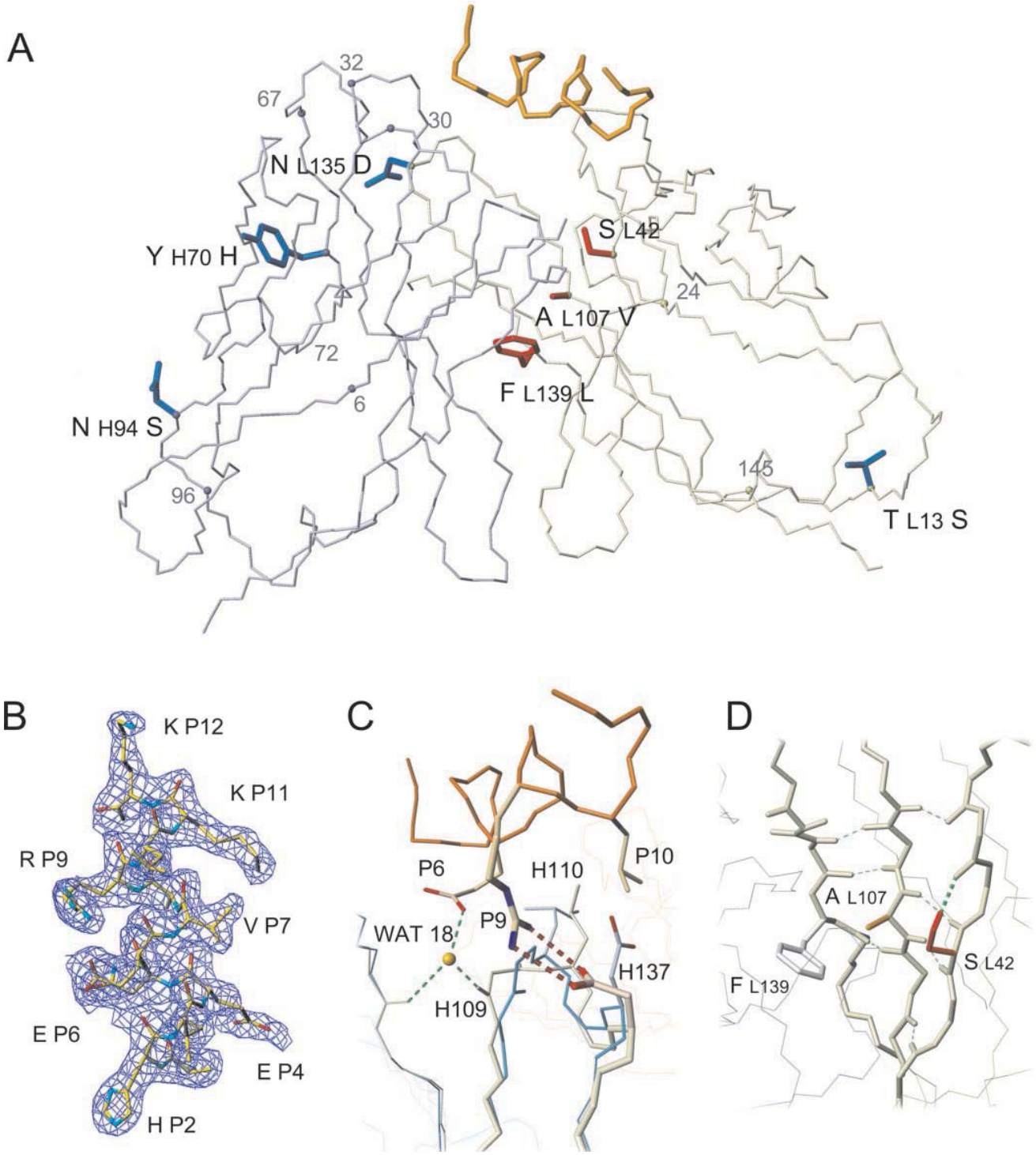


FIG. 4. **Overall structure of clone H6 in complex with the peptide and the location of the mutations.** In *A*, the backbone of the complex is shown. All mutations, which were found more than once in the directed evolution in any clone, are drawn with side chains including mutation L42(Asn → Ser), which led to clone C11L34 with 65-fold improved affinity compared with its likely murine progenitor. The Cα atoms of other mutations mentioned in the text are indicated by a *sphere* and a *number*. The light chain is drawn in *olive green*, the heavy chain is drawn in *light blue*, and the peptide is drawn in *orange*. The side chains of mutations lying in the interface of the two domains are shown in *red*. *B*, $2F_o - F_c$ electron density map contoured around the dodecapeptide. *C* shows the CDR H3 region overlaid in the antigen-bound and free state where the structural changes in the scFv due to antigen binding are significant. The antigen-bound form of the scFv fragment is drawn in *light orange*, and the free form is in *blue*, whereas the peptide is drawn in *orange*. Not all side chains are drawn. Arg-P9 establishes a strong interaction including two charged hydrogen bonds with Asp-H137 and thereby induces the conformational shift. The conformation of Arg-P9 may be locked by another weak hydrogen bond to Glu-P6. The greatest conformational change is undergone by Gly-H109, but there is a slight reorientation also by Leu-H110, which interacts hydrophobically with Leu-P10. The shift in the loop is stabilized by water molecule WAT18, which is hydrogen bonding to Glu-P6, Gly-H109, and Gly-H40. *D*, three mutations were lying close together in adjacent β-strands of the light chain facing the heavy chain. They may have an effect on the domain orientation and domain spacing. Residue Ser-L42 found in the previous selection is drawn in *red*. Ala-L107 (*orange*) and Phe-L139 (*gray*) were mutated to valine and leucine, respectively. The figures were prepared using the program MolMol (33).

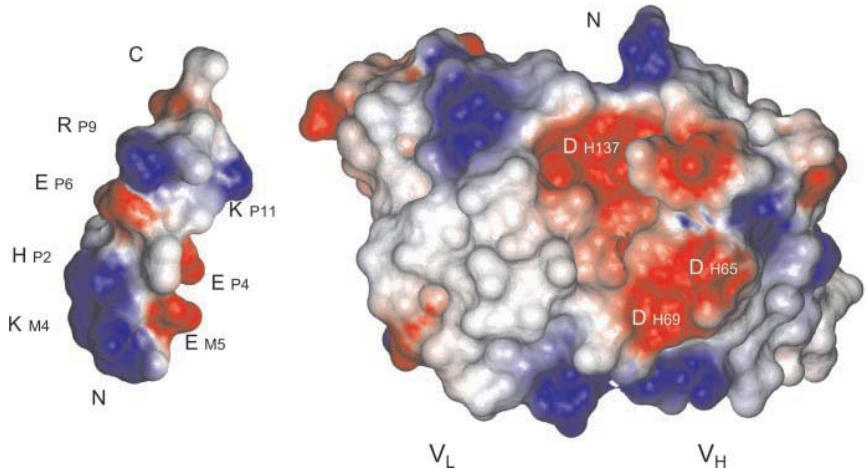


FIG. 5. **Electrostatic potential of the antibody with the antigen moved from the binding site.** A surface plot of the scFv and the antigen covering all amino acids potentially covering the binding pocket is shown. The six N-terminal amino acids DLPKQY of the peptide were modeled into the binding groove in a helical conformation. For better visualization, the antigen was removed from the scFv and rotated such that the binding face is visible. Folding the figure in the middle would restore the original orientation of the peptide.

mented electrostatically by the surface of the scFv (Fig. 5). Water molecules contribute strongly to the stabilization of the conformation of the peptide and are directly involved in interactions between the peptide and the antibody fragment. The conformational change in CDR H3 is probably favored by water molecule WAT18 (Fig. 4*B*).

A Structural Interpretation of the Evolution—Almost all the mutations that produced the high affinity binder are so-called "second sphere" mutations, which do not directly interact with the antigen and fulfill their beneficial effects via indirect interactions. The mutation L42(Asn → Ser), which leads to a 65-fold improved affinity of C11L34 relative to all its precursors and had been found in the original selection with ribosome display, is at least 6.6 Å away from the antigen. The light chain of C11L34 is of subtype λ. Interestingly murine germ line sequences of this subtype never carry a Ser at this position, whereas in murine light chains of κ-type, Asn and Ser side chains are common. In murine κ-type light chains, however, Asn side chains often make a hydrogen bond to residue H110 of V_H, stabilizing the interface, while Ser is rather seen in hapten binders, having a deep binding pocket, where the hydrogen bond is made directly to the hapten (*e.g.* Protein Data Bank entry 1C5C). In the case of C11L34, Ser-L42 plays an unusual role and forms a hydrogen bond to the backbone oxygen of Tyr-L40, stabilizing the conformation, which is typical for a backbone of λ-subtype. This stabilizing effect most likely reduces the flexibility of CDR L1, which is directly interacting with the peptide, through a hydrophobic contact between Tyr-L40 and Leu-P3. In addition, the space gained in the interface due to the smaller side chain may allow a more favorable domain orientation. A comparison of the scFv in the antigen-bound and unbound state reveals a domain rotation upon antigen binding and a reorientation of the CDR H3 loop, which brings CDR H3 1.5 Å closer to Ser-L42 than in the unbound state. In addition to Ser-L42, two other mutations, L107(Ala → Val) and L139(Phe → Leu), that showed up during the extensive off-rate selection lie in close proximity to Ser-L42 on adjacent β-strands of the light chain (Table II and Fig. 4*C*). These mutations are close to the pseudo-twofold axis of the scFv. It is likely that the interface mutations influence the relative domain orientation or domain spacing and thereby optimize the binding geometry.

The mutation H6(Glu → Gln) improved the affinity, by a factor of 2, to 20 pM when compared with clone C11L34. This residue is much too far away to interact with the antigen directly and must exert its beneficial effect by some long range interactions or "molecular shimming," influencing the orientation or flexibility of a loop or a domain. While the free scFv carries a glutamate at position H6, the structure of the scFv in complex with the peptide solved here is a variant carrying glutamine at this position. Position H6 was shown to define the conformation of the N-terminal part of the heavy chain (32). The conformation of the backbone in the free state is as it was expected for a glutamate. In the complex, however, this part of the structure is poorly resolved. Therefore, the occupancy of this area was set to 0. The poor density might indicate a conformational inhomogeneity or a higher than normal mobility, the reasons of which are unknown. The only mutation found during the selection that may interact directly with the antigen was introduced by site-directed mutagenesis during the library construction: L135(Asn → Asp). It most likely establishes a hydrogen bond to Lys-M4. However, this part of the peptide was not used for crystallization (see above). Most other mutations found during directed evolution are lying on the surface of the scFv fragment, and many may be neutral. However, some could also have indirect beneficial effects such as H30(Ser → Leu), which was found in the best affinity clone.

DISCUSSION

We have evolved a pool of single chain Fv antibody fragments against a peptide, which adopts a random coil in solution but a helical conformation in the complex, to achieve affinities of about 5 pM. We started from an already very tight peptide binder and generated a library by combining site-directed mutagenesis, DNA shuffling, and error-prone PCR. We applied directed evolution using ribosome display to improve the affinity a further 8-fold. Our result demonstrates that ribosome display is ideally suited for the identification of affinity-improving mutations and for the selection of binders under very stringent conditions even when already starting from picomolar affinities. In contradiction to previous concerns the stability of the ribosomal complex and the attached mRNA is not a limitation of the method.

One of the key results of the present study, analyzing the results from directed evolution crystallographically, is that only one mutation was making direct contact to the peptide. Three mutations were lying close together in the V_L/V_H interface. They may modulate the domain orientation, domain spacing, and the CDR loop flexibility. Many of the mutations found in the affinity-improved clones were found to lie on the surface

of the scFv not making direct contact to the antigen. Most likely these accumulated during the library generation and do not contribute to the improved affinities.

After the selection for a "first sphere" containing all important short range interactions, the overall geometry of the binding pocket was rearranged in a subtle manner. This was exclusively achieved by the mutation of key residues in the so-called second sphere, influencing the flexibility of binding loops and the orientation of domains rather than by changing interacting residues.

To our knowledge we have evolved the highest affinity reagent against a short unmodified peptide. Due to its very high affinity and the relatively small size of the antigen we expect the evolved scFv-peptide pair to be a powerful tool for biotechnological applications where tight binding to a tag is needed.

Acknowledgments—We thank Dr. Annemarie Honegger, Dr. Jozef Hanes, and Dr. Lutz Jermutus for helpful discussions and advice. We also thank Dr. David Zechel for critical reading of the manuscript.

REFERENCES

1. Amstutz, P., Forrer, P., Zahnd, C., and Plückthun, A. (2001) *Curr. Opin. Biotechnol.* **12,** 400–405
2. Hanes, J., and Plückthun, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4937–4942
3. Roberts, R. W., and Szostak, J. W. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 12297–12302
4. Smith, G. P. (1988) *Virology* **167,** 156–165
5. Fields, S., and Song, O. (1989) *Nature* **340,** 245–246
6. Mössner, E., Koch, H., and Plückthun, A. (2001) *J. Mol. Biol.* **308,** 115–122
7. Zaccolo, M., and Gherardi, E. (1999) *J. Mol. Biol.* **285,** 775–783
8. Stemmer, W. P. (1994) *Nature* **370,** 389–391
9. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A., and Virnekäs, B. (2000) *J. Mol. Biol.* **296,** 57–86
10. Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10701–10705
11. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., and Plückthun, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14130–14135
12. Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R., and Plückthun, A. (1997) *J. Immunol. Methods* **201,** 35–55
13. Bass, S., Gu, Q., and Christen, A. (1996) *J. Bacteriol.* **178,** 1154–1161
14. Adey, N. B., Stemmer, W. P. C., Kay, B. K. (1996) in *Phage Display of Peptides and Proteins* (Kay, B. K., Winter, J., and McCafferty, J., eds) pp. 280–292, Academic Press, Cambridge
15. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., and Plückthun, A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 75–80
16. Berger, C., Weber-Bornhauser, S., Eggenberger, J., Hanes, J., Plückthun, A., and Bosshard, H. R. (1999) *FEBS Lett.* **450,** 149–153
17. Ge, L., Knappik, A., Pack, P., Freund, C., and Plückthun, A. (1995) in *Antibody Engineering* (Borrebaeck, C. A. K., ed) pp. 229–266, Oxford University Press, New York
18. Auf der Maur, A., Zahnd, C., Fischer, F., Spinelli, S., Honegger, A., Cambillau, C., Escher, D., Plückthun, A., and Barberis, A. (2002) *J. Biol. Chem.* **277,** 45075–45085
19. Nieba, L., Krebber, A., and Plückthun, A. (1996) *Anal. Biochem.* **234,** 155–165
20. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497
21. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., eds) Vol. 276, pp. 307–326, Academic Press, New York
22. Collaborative Computational Project No. 4 (CCP4) (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
23. Rossman, M. G., and Blow, D. M. (1962) *Acta Crystallogr.* **15,** 24–31
24. Navaza, J. (1992) in *Molecular Replacement: Proceedings of the CCP4 Study Weekend* (Dodson, E. J., Gover, S., and Wolf, W., eds) pp. 87–90, Science and Engineering Research Council (SERC) Daresbury Laboratory, Warington, UK
25. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
26. Murshudov, G. N., Alexei, A. V., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
27. Roussel, A., and Cambillau, C. (1991) *Silicon Graphics Geometry Partners Directory*, pp. 81, Silicon Graphics Corp., Mountain View, CA
28. Karlsson, R. (1994) *Anal. Biochem.* **221,** 142–151
29. Schuck, P. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26,** 541–566
30. MacCallum, R. M., Martin, A. C., and Thornton, J. M. (1996) *J. Mol. Biol.* **262,** 732–745
31. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science* **254,** 539–544
32. Honegger, A., and Plückthun, A. (2001) *J. Mol. Biol.* **309,** 687–699
33. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14,** 51–55, 29–32
34. Jung, S., Spinelli, S., Schimmele, B., Honegger, A., Pugliese, L., Cambillau, C., and Plückthun, A. (2001) *J. Mol. Biol.* **309,** 701–716
35. Honegger, A., and Plückthun, A. (2001) *J. Mol. Biol.* **309,** 657–670
36. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S., and Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, National Institutes of Health Publication No. 91-3242, 5th ed., United States Department of Health and Human Services, Bethesda, MD

Curriculum Vitae

Name:	Amstutz
First Name:	Patrick
Date of Birth:	26th of March 1975
Place of Birth:	Zürich
Hometown:	Thalwil/ZH
Nationality:	Switzerland/USA

1994	"**Matura Typus B**", Kantonsschule Rämibühl, Zürich, Switzerland
1995-2000	**Studies of biology at the ETH Zürich** (molecular biology – biophysics, biotechnology, immunology, biochemistry, genetics)
1999-2000	**Diploma Thesis** in the department of biochemistry in the group of Prof. Dr. Andreas Plückthun. Title: "*In vitro* selection of catalytic activity with ribosome display"
2000-2004	**Ph.D. studies**: "Selecting Binders and Inhibitors from Ankyrin Repeat Protein Libraries" in the group of Prof. Dr. Andreas Plückthun, Department of Biochemistry, University of Zürich

Conferences, posters, oral presentations

1. Posters:

4th European Symposium of "The Protein Society" in Paris (France, April 2001) – *In vitro* Selection of Catalytic Activity with Ribosome Display

INPEC Meeting in Cambridge (UK, September 2002) – Selection of Enzyme Inhibitors from a Designed Repeat Protein Library

2nd Symposium of the NCCR Structural Biology in Zürich (Switzerland, November 2002) – Selection of Enzyme Inhibitors from a Designed Repeat Protein Library

Drug Discovery Technology Europe 2004, IBC Meeting in London (UK, March 2004) - Designed Ankyrin Repeat Proteins: A True Alternative to Antibodies

2. Oral presentations:

Invited Talk at Cambridge Antibody Technologies (UK, November 2000) - Functional selection of ribosome-displayed β-lactamase

Protein Microarray Technology, ESF Exploratory Workshop in Tübingen (Germany, October 2001) - Directed Evolution: selecting high affinity and high stability *in vitro*

Invited Talk at Cambridge Antibody Technologies in Cambridge (UK, September 2002) - Repeat protein libraries as a source for intracellular binders and enzyme inhibitors

Protein Engineering & Discovery, IBC Meeting in San Francisco (USA, March 2003) – Synthetic Repeat Proteins

Invited Talk at Genentech in San Francisco (USA, March 2003) - Selection of Binders and Inhibitors from a Designed Repeat Protein Library

Invited Talk at UC Berkeley (USA, March 2003) - Selection of Binders and Inhibitors from a Designed Repeat Protein Library

Drug Discovery Technology Europe 2004, IBC Meeting in London (UK, March 2004) - Designed Ankyrin Repeat Proteins: A True Alternative to Antibodies

Invited Talk at Cambridge Antibody Technologies in Cambridge (UK, March 2004) - Selection of Binders and Inhibitors from a Designed Repeat Protein Library

Invited Talk at Apoxis in Lausanne (Switzerland, April 2004) - Selection of Binders and Inhibitors from a Designed Repeat Protein Library